Statistics Canada
Canadian Centre for Justice Statistics

Statistique Canada
Centre canadien de la statistique juridique

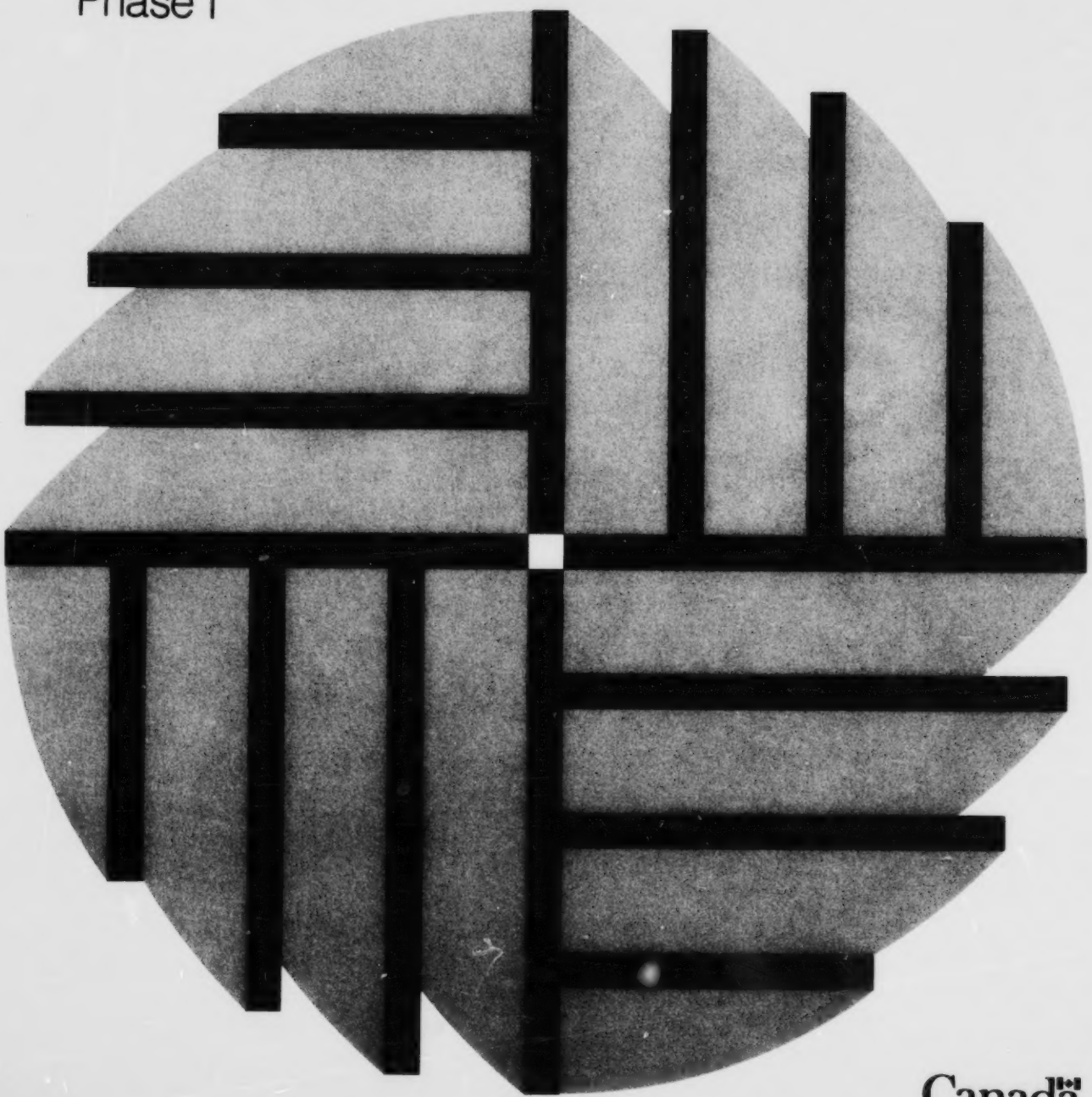
R-051B

(March 1985)

c.2

Review of Non-fingerprint Supported Identification Technology

Phase I



Canada

REVIEW OF NON-FINGERPRINT SUPPORTED IDENTIFICATION TECHNOLOGY

PHASE I

CONDUCTED FOR THE

CANADIAN CENTRE FOR JUSTICE STATISTICS

AND

FUNDED BY SOLICITOR GENERAL CANADA
THROUGH THE YOUNG OFFENDERS ACT SYSTEMS DEVELOPMENT
CONTRIBUTION PROGRAM

PREPARED BY

SYNERLOGIC INC.
(FORMERLY BAILEY & ROSE LTD.)

MARCH 1, 1985

REVIEW OF NON-FINGERPRINT SUPPORTED IDENTIFICATION TECHNOLOGY

PHASE 1

TABLE OF CONTENTS

		Page
1.	**Executive Summary**	1
2.	**Introduction**	2
	2.1 Background	2
	2.2 Objectives and Activities	3
3.	**Technology Requirements**	4
4.	**The Identification Question**	6
5.	**Technology Summaries**	10
	5.0 Introduction	10
	5.1 British Columbia Provincial Case File System	10
	5.2 Alberta Court Automated Project	12
	5.3 Alberta Corrections Management Information System	13
	5.4 Saskatchewan Corrections	14
	5.5 Ontario Corrections	15
	5.6 New Brunswick Justice Information System	16
	5.7 Nova Scotia Social Services	18
	5.8 Royal Canadian Mounted Police	19
	5.8.1 The Royal Canadian Mounted Police (RCMP) PERSONS System	19
	5.8.2 Additional Royal Canadian Mounted Police (RCMP) Technologies	22
	5.9 Utah Juvenile Justice System	22
	5.10 Maricopa County Juvenile On-line Tracking System	24
	5.11 Memphis Juvenile Justice System - A JISRA Implementation	25
	5.12 Toronto Credits Limited - A Private Sector Application	27
	5.13 GIRLS - A Record Linkage Application	29
	5.14 Summary Table of Characteristics	32
6.	**Technology Details - The Short List**	35
	6.0 Introduction	35
	6.1 Ontario Corrections	35
	6.2 Nova Scotia Social Services	37
	6.3 Royal Canadian Mounted Police	39
	6.4 Utah Juvenile Justice System	42
	6.5 Maricopa County Juvenile On-line Tracking System	47
	Appendix A - List of Contacts	48

1. EXECUTIVE SUMMARY

This report presents the findings of Phase 1 of the "Review of Non-Fingerprint Supported Identification Technology", a project of the Technical Assistance Directorate of the Canadian Centre for Justice Statistics. The overall project was initiated as a result of interest on the part of a number of jurisdictions in developing integrated information systems to facilitate the administration of the Young Offenders Act (YOA). An integrated information system, where the subject is an individual, must be able to uniquely identify that individual within the system. Fingerprints provide a reliable and positive identifier but can only be used for a small proportion of offenders.

The alternate methods currently being used to identify individuals were reviewed in this phase of the project. The results of this review only identified namesearching (the use of an individual's name and optional secondary identifiers such as date of birth and sex) as an alternative to fingerprint identification.

This report provides a technical overview of the various implementations of namesearching technology. Chapter 4, The Identification Question, may be of interest to readers who wish to familiarize themselves with the basic concepts of namesearching and the issue of identification. Chapter 5 provides an overview of thirteen information systems which employ namesearching technology and Chapter 6 provides more detailed technical descriptions of the specific techniques used in five systems.

Prior to publication, this report was distributed throughout Canada for comments and review. Two options were presented regarding future research into namesearching technology:

1) The provinces and territories could use the report as an informational document and independently investigate the suitability of the technologies for their own systems development efforts. No further action would be taken by the Canadian Centre for Justice Statistics.

2) The Canadian Centre for Justice Statistics would continue its investigation into the five technologies detailed in Chapter 6. The report to be produced would be a technical document suitable for use in detailed systems design.

The jurisdictional response indicated that the undertaking of a Phase 2 study by the Canadian Centre for Justice Statistics was desirable. The second report will incorporate the following information:

- the interfaces between the identification function and the overall functional components of the information system;

- the design details of the identification function;

- the transferability and/or flexibility of the program code or the programming language;

- the problems associated with the technologies including the risks and success rates as identified by the experts associated with these technologies; and,

- the checks and balances in place to minimize the identified risks.

2. INTRODUCTION

2.1 Background

The proclamation of the YOA has resulted in an increased interest throughout the provinces in integrated justice information systems. The record keeping and destruction provisions of the YOA, in particular, have brought to light the need to integrate information concerning an individual regardless of the justice system sector involved and the time or location of the event.

A key issue in an integrated information system, where the subject is an individual, is the identification of that individual. In a sector specific system that is concerned with an isolated event, the identification of an individual poses no significant problem, as there is a natural progression of the individual through the system. A successful integrated system, however, must be capable of uniquely identifying an individual regardless of the sector, time or location. An appropriate identification system with inherent checks and balances requires considerable time and care to establish.

Several organizations in Canada and the United States have in place information systems in which the identification of an individual is accomplished without the aid of fingerprints. The justice community can benefit from the experience of those responsible for these systems when developing information systems which will meet the requirements of the YOA.

2.2 Objectives and Activities

The objective of this project, as stated in the project proposal is:

'To learn from the experience of others the advantages and disadvantages of non-fingerprint technology and to quantify the risks'.

This report provides an overview of existing identification technologies not relying on fingerprints which might be appropriate for use in information systems relating to juveniles.

In completing Phase 1 of this study, the following topics have been addressed.

- what identification systems exist and where?
- what are the characteristics of these systems?

- what is the functional relationship of the identification component to the overall operating functions of the information system?

The activities that comprised this project include:

- interviews with Canadian Centre for Justice Statistics personnel;
- interviews (primarily by telephone) with persons cognizant of identification systems in Canada and the United States;
- evaluation of the identification issue as it exists today;
- documentation of findings regarding thirteen existing systems; and
- evaluation and further documentation of a representative group of five existing systems.

3. TECHNOLOGY REQUIREMENTS

In this phase of the study, information systems and their associated identification technologies are not specifically being evaluated to assess their suitability to meet the needs of provincial information systems. Subsequent phases of the project would deal with their adaptability to provincial or territorial systems, risk factors and success rates. Another study also undertaken by the Canadian Centre for Justice Statistics, the Preliminary Evaluation of Software Packages related to the Young Offenders Act, should complement this analysis of identification techniques. This phase will review existing identification technologies within the framework of the overall system context and report on a representative sample of these systems, providing a base for further study.

To accomplish this task, it is necessary to identify a general requirements list to ensure that the systems reviewed during this phase will indeed provide a base for further study. The following requirements were used as a general guideline when evaluating the suitability of systems for inclusion in this report.

. The technologies must be able to store and retrieve information concerning individuals through means other than fingerprints.

. The technologies must have tracking capability. Identification of the individual must be possible regardless of the time or location of the event. The assumption can not be made that the collection and coordination of data will be centrally tracked.

. The identification technology must, in general terms, be appropriate for use in information systems related to juveniles. Technologies relying heavily on characteristics not normally attributable to juveniles may not be appropriate. Technologies, for example, which use the Social Insurance number (SIN) or the driver's licence number, may not be appropriate for use in juvenile identification systems.

. The information system should be designed for operational as well as statistical use. A pure statistical system may have a greater risk tolerance in terms of tracking and data integrity.

The mandate of this study is to provide a broad representation of the existing identification technologies upon which to base further analysis. Various systems across Canada and the United States have been reviewed and documented and a "short list" of five systems has been selected for further study. This "short list" of systems was chosen to provide the best possible representation of technologies. In light of this approach the following factors were taken into consideration when selecting the "short list".

. The Justice System includes the Courts, Corrections and Law Enforcement. A representative technology should be provided for each sector.

. Preliminary research into the question of identification and the existing systems indicated that different approaches have been adopted for use (see the Section "The Identification Question"). Applications of these different approaches should be included for further study.

. The "short list" may include systems operating either in Canada or the United States. Although the U.S. systems included in this study may be well-developed in the overall system context, it became apparent that the identification component of some Canadian systems is technologically comparable to that of the United States.

4. THE IDENTIFICATION QUESTION

Accurate identification is essential to the success of any individual-based information system. In the area of criminal justice records, the identification function must be able to integrate information concerning an individual regardless of the sector being served or the time or location of the event. Considerable research has been applied to the development of automated and semi-automated techniques for the identification of serious offenders through the use of fingerprints. Classification based on fingerprints provides a reliable, positive means of identification for only a small portion of offenders. The less serious offences must still be related to the individual. This is accomplished by a technique called namesearching using a person's name and personal attribute identifiers such as sex or date of birth.

Namesearching systems based on personal attributes do not generally rely on these personal attributes as the primary unique identifying elements upon which record linkages are based within the computerized files. Coded identifiers, often arbitrarily assigned numbers, may be the unique keys used for storing records, and in some cases, retrieving information concerning an individual. The coded identifier, if known by the operator, may be used in conjunction with the personal attribute identifiers to perform searches. The personal attribute identifiers are used primarily to:

- establish prior contact with the system (does the computer system already have information concerning the individual) by providing lists of individuals on file who have matching attributes to the subject individual. These lists are used by the operator to judgementally determine whether or not the subject individual has previously had any information entered on the system; and

- provide a basis for efficient automated storage organization and retrieval access.

A report prepared for the Federal-Provincial Committee on Justice Information in April 1976[1] examined the question of identifiers and reviewed the following alternatives:

- Local Administrative Number - a number assigned at the local level;
- FPS Number - the fingerprint serial number, available only for serious offenders;
- Government-assigned Number - included in this category are Social Insurance Number, Driver's Licence Number; and,
- Name and Personal Statistics.

The alternative considered in the 1976 study to be the most suitable as a prime identifier was an individual's name. The study acknowledged that the name would only narrow identification to a choice. The assumption was also made that the individual would willingly divulge the information.

In March 1975, Search Group Inc. of Sacramento, California published a report which provided an overview of the state of the art in criminal justice identification. The identification systems included in this

---

[1] The Needs for Data Linkages in the Corrections Component of the Criminal Justice System, Report Prepared for the Federal-Provincial Committee on Justice Information by the Task Force on Identifiers, April 1976.

report were Automated Namesearch systems and Automated or Semi-automated Fingerprint Search and Taking techniques. The report states that "all necessary research in the namesearch area has been completed"[2]

The review of information systems undertaken in this study has revealed that existing criminal justice information systems are continuing to rely on either fingerprint-based or namesearch techniques to identify individuals.

A more recent study, undertaken in 1982 by the U.S.-based Office of Technology Assessment, assessed the alternatives for a national computerized criminal system. In reviewing the current status of criminal history record systems, the report states: 'Criminal justice practitioners believe that, at present, fingerprints are the only reliable and consistent basis for positive identification. The exchange of records based on names alone results in a high percentage of errors due to the frequent use of aliases and similarities among many common surnames'.[3] Considerable effort has been applied to the development of techniques to minimize risks associated with the spelling and similarities of surnames. These techniques are referred to as namesearching in the remainder of this report. The information systems included in this report are all using some type of namesearch technique.

The remainder of this section will familiarize the reader with the basic approaches applied in namesearching and the terms used to describe the methodologies employed by the systems. It must be remembered that the purpose of namesearching is to determine whether the individual has records on the computer system. Any subsequent actions are entirely system-dependent.

The term namesearch could encompass the following processes:

---

[2] Search Group, Inc., Search Group Masterplan for Identification Systems Upgrade, March 1975.

[3] An assessment of Alternatives for a National Computerized Criminal History System, a project of the Office of Technology Assessment, Congress of the United States, October, 1982.

1) Determination of the group of records to be searched.

In larger systems, it would not be desirable to physically search every record stored to determine matches. A Name Grouping technique is often employed to segment the file in order to reduce the number of physical record searches.

2) Searching the group of records.

In this step, both the name and any additional or secondary identifiers are used to determine possible matches between the subject individual and file records. There are, generally, two types of search methods.

An Exclusion Search will exclude from further consideration any records which do not match any search argument. In this case, an exact match or a match within a certain tolerance level is attempted using each identifier in order. Any one mismatch is considered a miss.

A Scoring Search considers all of the match identifiers when determining a hit or a miss. Each element is assigned a score depending on a match, a mismatch, or a missing element and the sum of those scores indicates the probability of a record match.

3) Listing

The results of the Grouping and Searching steps are listed. Generally, all of the matches are listed if an Exclusion Search was used. If a Scoring Search was used, two types of lists can be returned, a list of records in order of their score or a list of all records that scored above a predetermined threshold score.

4) Verification

The final identification decision is usually a manual procedure.

Name Grouping is often accomplished by grouping together records in which the last name sounds similar (known as a phonetic routine or algorithm). The various phonetic routines currently being used are sometimes referred to as soundex routines. The original soundex routine is 'Russell Soundex', a system developed for manual searching of name indexes. Two widely-used phonetic routines in North America are the RCMP FIND Code and NYSIIS, the New York State Information and Intelligence System code.

5. TECHNOLOGY SUMMARIES

5.0 Introduction

This chapter provides an overview of thirteen information systems that have a functional requirement to identify individuals without the aid of fingerprints. Various systems in Canada and the United States were identified primarily by the 'Inventory of Automated Justice Systems in Canada', a project of the Technical Assistance Directorate of the Canadian Centre for Justice Statistics, and through contact referrals. The thirteen systems chosen for inclusion in this chapter are by no means meant to represent a complete review of all of the information systems currently performing identification functions but they were chosen to provide as broad a base as possible for further study.

A section is included for each of the thirteen systems, describing the purpose and use of the system and the functional components. The basic characteristics of the systems, in terms of their identification function, are then summarized in tabular form in 5.14.

5.1 British Columbia Provincial Case File System (PCFS)

PCFS is an operational tracking system maintaining personal, history, court and location data for both active and inactive clients of the British Columbia Corrections Branch.

The operational functions supported are:

- Client Identification - (used for prior contact searches);
- Client Information;
- Client History; and,
- Client Tracking

In addition to the operational functions, the system provides a statistical analysis function to monitor Corrections Services utilization and to support long-range planning.

PCFS is a Head Office system with a minimum of one access terminal located in each of the five regions of the province. Summary data concerning all individuals who have had contact with British Columbia Corrections over the last ten years is maintained by the system. More detailed client information is maintained at the institutional level by the Local Institution System (LIS). An interface exists between PCFS and LIS to permit assignment of the Corrections Service Number. The CS # serves as the unique identifier linking all records for a particular individual.

The client identification function of PCFS provides three search options:

1) Exact Name - using surname, given name and year of birth. A variance of two years is allowed in year of birth.

2) Partial Name - using truncated surname, given initial and year of birth. Any number of letters may be truncated in the surname.

3) Soundex - A phonetic search using a separate data base. The Soundex option originally had three levels of selectivity. The highest selectivity level resulted in the smallest group of possible matches, the lowest in the largest group of possible matches.

The system has been modified to allow access to the highest selectivity level, only, as too many possible respondents were being returned using the other options.

The client identification function at the institutional level is restricted to the exact and partial name options, only.

PCFS operates on an IBM 3081 and uses IMS ADF version 1.3.

5.2 Alberta Court Automated Project (CAP)

CAP is a province-wide system serving 3 levels of adult courts in Alberta. Data is entered from 26 Provincial Court locations, 13 Court of Queen's Bench locations, and 2 Appeal Courts. CAP is a comprehensive court system covering most court functions including:

- Court scheduling
- Case tracking
- Warrants
- Financial control, including bail, restitutions and fines
- Document generation (summons, warrant, etc.)
- Tickets
- Accused History

The original system handled only the ticket function. Phase 2, the current system, has been fully operational for approximately one year. Enquiry and data entry is on-line but the update function is handled by a nightly batch run. The system currently maintains approximately 8,000,000 records.

The accused/history enquiry function is accessed to initiate a search for all cases associated with an individual. The identifiers that may be used to conduct a search are: surname, sex, date of birth, court code and/or driver's licence number.

A comprehensive phonetic code, similar to that used by the RCMP, but less finely tuned, is used for the surname but no scoring capabilities are available.

CAP operates on an IBM 3033 located at the Central Computing Centre in Edmonton, using TOTAL/ENVIRON.

5.3 Alberta Corrections Management Information System (COMIS)

COMIS was implemented in 1979 to serve both the operational and statistical analysis requirements of the Alberta remand and corrections institutions. The operational components include:

Inmate Tracking - monitors inmate movements and prepares court and transfer lists.

Sentence Calculation - calculates most variations of sentence dispositions.

Client Tracking - searches files to determine if the individual has previously been assigned an offender identification number.

A research file containing data back to 1977 serves as a basis for the statistical analysis function. Five years of historical data is available on-line.

Current development includes:

- conversion from one data base package to another (TOTAL/ENVIRON TO ADABAS/NATURAL) to provide more flexibility;
- addition of probation information;
- interface with the Canadian Police Information Centre (CPIC) to allow police interrogation of COMIS files; and,

- interface with the Court Automated Project (CAP) of the Attorney-General to provide weekly transfer from COMIS to CAP of information concerning discharges.

Two remand centres have operational mini-based systems. Other institutions are on-line to the central mainframe to perform namesearches and record admissions and releases.

The identification function of COMIS is the key element of the inmate tracking component. A main objective of COMIS is to accumulate complete case histories under a unique inmate file number. With a recidivism rate of over 35%, a technique must be available to determine whether the inmate is a returnee. The admission procedure originally performed by the mainframe used a basic alphabetic match on the first four characters of the surname. This technique had a number of deficiencies which necessitated a major upgrade in 1983 to the phonetic namesearch technique currently in use. The upgraded technique is virtually identical to the Russell Soundex system (see 6.6). The primary identifier is surname and secondary identifiers include first name, birth date (or age), social insurance number, height and weight. The COMIS search technique does not have probability scoring capability.

The head office system runs on an IBM 3033 and currently uses TOTAL/ ENVIRON. The operational institution systems use Data General Eclipse 350's.

5.4 Saskatchewan Corrections

The adult correctional community in Saskatchewan is served by a Head Office batch system supporting research and statistics and an on-line system supporting the operational functions of the three largest correctional centres.

The Head Office system, Correctional History System, was developed in 1976 and maintains correctional and probation information for inmates of all of the Saskatchewan institutions. Data is received from the manual institution systems in the form of hard-copy documents. The three larger, automated systems generate card images of current inmate data on weekly key-edit tape to be used as updates to the Head Office system. There are no automated interfaces between the Correctional History System and the institutional systems.

The Correctional Centre Information System (CCIS) was recently developed as a mini-based operational system for the three largest centres. Four terminals, one in each of the centres and one in head office are connected to an IBM model 8100 mini-computer. Full inmate records are maintained for current cases, only, but skeleton records are maintained for two year's prior data. It is planned to expand the skeleton records to cover five years of data. The information that is maintained by the CCIS is duplicated on the Head Office batch system. The system matches on the first five characters of surname, given name initials, and month, day or year of birth date.

5.5 Ontario Corrections

The operational and statistical information requirements of the Ontario Ministry of Correctional Services are supported by automated systems at both the Head Office and Institution level.

The batch Adult Information System (AIS) has been in existence for over 8 years and during that time has collected information on the approximately 300,000 individuals who have been involved with the provincial correctional system. AIS is a Head Office system that provides a base of data for statistical analysis, probation and parole management, sentence calculation audit and client history profiles. Inmate/client data is forwarded to AIS by all institutions and probation and parole officers.

The large, computerized institutions transmit data directly to the mainframe on a daily basis. The other institutions and probation and parole officers submit data to the main office in the form of hard-copy documentation. In either case, the AIS main file is updated in batch mode on a weekly basis.

The on-line Rapid Enquiry System (RES) provides, as its primary objective, the ability to determine whether the individual has been in the system before. The RES on-line data base is an extract of the data contained on the Adult Institution System (AIS). Summary data is maintained by RES for every individual who has been in contact with the correctional system. RES serves as an identification enquiry system, only. Complete client/inmate profile data is a function of AIS.

The Local Institution System (LIS) is a mini-based system serving the operational requirements at the institution level. Currently seven locations are operational using LIS. The remaining institutions are manual systems.

Long-term development plans for these three correctional systems include technological upgrade and system integration of the functions supported by the three individual systems.

The identification function, handled exclusively by RES, uses name as the primary identifier and sex, birth date or partial birth date as the secondary identifiers. A well-developed phonetic search routine is used to minimize the risk of spelling mismatches on surname.

AIS operates on an IBM 3033, RES on a Data General MV8000 and LIS on Data General 5140's. Dial-up lines are used by the local institutions to transmit data to AIS.

5.6 New Brunswick Justice Information System (JIS)

JIS is a case-management system dealing with indictable and summary conviction offences. It allows for the initial creation of a case record and the addition of information to the record as case proceedings take place. JIS handles an annual caseload of approximately 50,000.

The cornerstone function of JIS, case management, comprises the following elements:

- presentation of accused tombstone information;
- dates of court appearances; and,
- identification of police officers, witnesses, judges, attorneys and prosecutors.

Other functions of JIS include:

Fine Control - A financial component to facilitate the control of payments and outstanding fines.

Bail Control - a financial component to facilitate the control of payments.

Restitution Control - a financial component supporting the recording, collecting, compensation, and reconciliation of payments.

Court Scheduling - court schedules are prepared up to two weeks in advance and are used to notify judges, prosecutors and police officers.

Warrant Control - allows for the registration and follow-up of warrants.

Reporting - statistical report formats are available to support the case management function.

Control Files

Management - a series of control files are used for editing purposes to identify judges, prosecutors, police officers, witnesses, criminal code sections, bylaw code numbers and provincial statutes. The defendant file is the only tombstone data file.

JIS has been operational since 1982. It is anticipated that prosecutors will rely on the system for tracking as it ages.

JIS uses the NYSIIS code developed for the New York State Police in its phonetic namesearch process. The same phonetic routine is applied to both the first and last name. There are no secondary identifiers used. All cases associated with the coded name are returned rather than any individual identifying data.

JIS is a modified version of PROMIS (Prosecutor Management Information System), a U.S. software product developed by INSLAW Inc. of Washington, D.C.

Sequential and Indexed Files are used in an on-line environment on a UNIVAC 9030 computer.

5.7 Nova Scotia Social Services Client Index System

The Department of Social Services provide a wide range of social services to families, children and the disadvantaged. The Department must ensure that not only are individual programs delivered in the most beneficial manner, but that the best mix of services is being delivered in an integrated and coordinated manner.

The Client Index System is the component of the Management Information System that ties together all programs, helping the Department to achieve this goal. The Client Index System performs the following functions:

- maintains a record of all clients;
- provides a mechanism for prior contact checking;
- provides housekeeping and fraud detection routines;
- contains non-program specific or common client data;
- facilitates operational sharing and exchange of information between programs; and,
- provides regular management reports tracking clients over time and across programs.

The use of the Client Index System as a front-end prior contact check is straightforward. Every time an individual comes into contact with the department, a check is made to see if the individual is known to the system. The primary identifiers are sex, date of birth, surname, first name or initial. If the individual is unknown, a record is established on the Client Index which contains personal identification data. A record of program involvement is recorded separately.

The Client Index system in Nova Scotia operates on an IBM 4341 with both on-line and batch mode capabilities. The Client Index uses a VSAM file structure.

This system originated in Nova Scotia using a Client Index System in Quebec as a model and has an extensive transfer history. Versions of this social service system are being used in Saskatchewan, Alberta, Alaska, North Dakota, and Idaho. Alaska and North Dakota have replaced the VSAM file structure with ADABAS and upgraded the system to include income security programs. Saskatchewan plans to convert the ADABAS but is currently using VSAM.

5.8 Royal Canadian Mounted Police (RCMP)

5.8.1 The Royal Canadian Mounted Police (RCMP) PERSONS System

PERSONS is an on-line operational system used to identify individuals such as wanted or missing persons, parolees or restricted persons. PERSONS provides a means of identifying these individuals, points to the case and specifies the reason that these individuals are of interest to the police community. This system is directly accessible by law enforcement officers exclusively but indirect access to the information is available to other authorized agencies (Courts, Corrections). Law enforcement officers routinely access PERSONS to determine if an individual is of any interest to the police community. For example, an individual who has been stopped for speeding will be asked, among other things, his or her name. The PERSONS system will then be searched to determined whether this name (qualified perhaps by other identifiers) matches any individuals identified to the system. The results of the search could indicate that a record exists on PERSONS for an individual who is very likely the same individual as the one stopped for speeding. This information enables the officer to determine the action to be taken.

The techniques developed by the RCMP for identifying individuals are widely used across Canada and the United States. The Alberta Corrections Management Information System, the Edmonton City Police, National Archives, Canada Employment and Immigration, the Provinces' Registered Owners systems and the Central Intelligence Agency in the United States all use versions of the RCMP technique. The ultimate application of this approach, however, is used in the PERSONS System.

The RCMP Criminal Records Synopsis (CRS) System also uses this technique to identify individuals. CRS acts as an index, pointing to full criminal history records on the Criminal Records System (CR2) which is a fingerprint-supported system.

In PERSONS, twelve personal descriptor data elements are used as keys when searching for an individual. The surname is a mandatory search key and is used to phonetically segment a group of records for searching. Unlike most phonetic or soundex name grouping techniques, the RCMP technique allows for alternate phonetic codes to be applied through an extra cross-referencing step. A comprehensive scoring algorithm is then applied to each personal descriptor secondary element in each record of the selected segment to determine the probability or likelihood that the record on the PERSONS system matches the subject individual. The results of the search are then returned in highest probability sequence. At this point, the system operator judgementally determines whether the subject individual is present on the PERSONS system.

RCMP PERSONS uses an internally developed data base system called PRISM and operates on an IBM 3033 N, with 16 megabytes of storage. The file organization method of PRISM was designed to suit the inherent characteristics of a criminal file.
The RCMP namesearch technique is referred to as the FIND code and encompasses three basic components:

1) PHONETIC coding of the surname to determine the segment to be searched.

2) XREF or cross-reference phonetic coding of the surname to determine alternate segments to be searched.

3) SCORING which assigns weights to the key data elements to determine the match probability.

Although other systems are currently using versions of the FIND code, they vary in the degree to which the techniques are applied.

5.8.2 Additional Royal Canadian Mounted Police (RCMP) Technologies

In addition to the individual identification systems developed by the RCMP is an intelligence system, the Automated Criminal Intelligence Information System (ACIIS). ACIIS is the only system identified during research for this report that does not necessarily rely on a name as the primary identifier. Weighted or exact match searches are performed using up to sixty different identifying keys, including personal characteristics such as marks, scars, deformities and body features. ACIIS is worthy of mention, but was not considered appropriate for further study, in that it is an accumulation of data (as opposed to an individual identification system), requiring specialized intelligence analysts with extensive training in the operation and use of the system.

Advanced technological development by the RCMP includes a new application which will rely on facial characteristics as identifiers. Computerized 'mug shots' may become a key identifier in both the Criminal Intelligence System and the Criminal Records Synopsis System.

5.9 Utah Juvenile Justice System (Profile II)

Utah's Juvenile Court has been using computers to collect information and process cases since 1967. The automated system has evolved from a card-based collection of data from the court intake form for a juvenile index, rap sheet and statistics to the present statewide terminal network linking all elements of the juvenile justice system.

Profile II is described as an operational record system, designed to assist the Juvenile Justice System accomplish their tasks which include:

1) Initial Case Processing - The system conducts prior contact checks, captures both statistical and processing data and produces an updated rap sheet;

2) The Intake Process - The rap sheet produced as a by-product of initial processing assists the intake officer in determining the course of action to be followed. The system is able to produce the petition, summons, court docket, fine notices or appearance schedule;

3) Tracking Court Orders - A summary of court actions is recorded on the Court Docket during each hearing, which is subsequently posted to each record;

4) Tracking Physical Custody - The Detention Centres and Youth Corrections use Profile II to book youth in and out of their facilities and maintain daily population lists; and,

5) Special Features - Key special features include access to adult criminal history records, Driver's Licence Records, Motor Vehicle Registration records, ad hoc reporting capabilities, cheque printing and accounting, and access to the National Law Enforcement Telecommunications Network (NLETS).

The identification of juveniles during Initial Case Processing uses a technique developed by SEARCH Group, Inc. of Sacramento, California, the National Consortium for Justice Information and Statistics.

MICRONYM, a system for identification processing, is a comprehensive package providing for the management and operation of document processing and namesearch within a state identification bureau. MICRONYM uses the name grouping technique developed in 1970 for the New York State Information and Intelligence System (NYSIIS). A scoring search measures the probability that the file record represents an identity match with the source document subject.

Profile II is maintained on two AMDAHL computers (V6 & V8) using ADABAS/COMPLETE, a database teleprocessing software package.

Courier terminals (IBM 3270 compatible) are located at 14 juvenile court locations, 6 youth corrections sites, 31 police dispatching points and 6 adult probation and parole centres statewide. Most programs are written in COBOL. NATURAL is used to develop ad hoc, specialized programs to access the data.

5.10 Maricopa County Juvenile On-Line Tracking System

The Juvenile On-Line Tracking System (JOLTS) provides an automated data system for the Maricopa County Juvenile Court, a branch of the Arizona State Superior Court. The prime geographic area of jurisdiction covers a total population of approximately 1.4 million people, 428,475 of which are under the age of eighteen.

JOLTS was originally designed to monitor delinquent and incorrigible cases replacing a burdened manual card system. Three years in the planning, JOLTS met its three main objectives:

1) to provide rapid and accurate information feedback concerning clients to probation officers and other appropriate criminal justice agencies;

2) to provide reports for management; and,

3) to increase research capability.

Enhancements to the original system have included case tracking, a traffic system, and court calendaring.

The identification component of JOLTS uses a soundex filing system that contains both the names and file numbers of every juvenile on the database. The primary identifier is juvenile surname and the optional secondary identifiers are first name, sex and date of birth. All juveniles that fall within the name group are considered "hits" and will appear on the screen. JOLTS employs no probability scoring technique per se but the "hits" are sequenced by the degree of exact matches to aid in the judgement process. A quality control report is run monthly to identify possible duplicates.

JOLTS operates on a HONEYWELL 66/40 in both batch and on-line modes. A recent assessment of the JOLTS system has recommended conversion to more current hardware.

5.11 Memphis Juvenile System - A JISRA Application

The Juvenile System in Memphis, Tennessee is the most recent mainframe implementation of the JISRA (Juvenile Information System & Records Access) model. JISRA is a project of the National Council of Juvenile and Family Court Judges, an U.S.-based organization. JISRA supports management, outside reporting and special research needs of juvenile courts and is described as an on-line event and person tracking computer system designed for transfer.

JISRA can be transferred to any IBM compatible equipment and any microprocessor supporting the OASIS operating system. Since the JISRA system was designed for use in a variety of environments, it has been implemented as a series of modules which can be selected or revised to reflect the functional needs of each court.

The following core JISRA modules, available on all versions, may function differently in each implementation:

1) Security	Operator security is a function of both mainframe and microcomputer environment. Terminal security is available for mainframes, only.
2) System Edits	Data field editing is table-driven.
3) Juvenile	To gather and maintain both identification and environmental data concerning a juvenile.
4) Namesearch	To establish prior contact.
5) Case Summary	Provides a juvenile history including dates, and referral dispositions.
6) Referral/ Intake	Recording of basic information describing each referral.
7) Calendar	Produces court schedules.
8) Hearing/ Disposition	Records hearing, incident and disposition data.
9) Detention	Records residential care data.
10) Statistics	To measure effectiveness, indicate trends, and project future needs.
11) Report Generation	Pre-defined report formats.
12) Alias	Several alternate names can be retained for the same juvenile.

13) Adults — To gather information about the adults associated with the juvenile.

14) Family — To provide a descriptive list of all family members.

15) Other Person/ Witness — To identify persons associated with a referral or complaint.

The namesearch module associated with the Memphis mainframe application applies a phonetic routine to the juvenile's name to determine prior contact. The system returns a list, in alphabetic sequence, of all juveniles on file having the same or similar names (real or alias).

There are no secondary identifiers used in the JISRA model. The identifying data elements that appear on the screen to aid in the determination of prior contact are first and last name, aliases, sex, date of birth and race.

The Memphis application runs on an IBM model 4331.

5.12 Toronto Credits Limited - A Private Sector Application

The requirement for identification of individuals exists in a variety of organizations and agencies. The needs of the justice community and the social services community are similar in that individuals must be tracked through time and from agency to agency. The aims and data collection procedures of private sector organizations may be quite dissimilar from those of the public sector but the problem of identification is a common concern. For example, credit bureaus maintain large files of individual-based data against which enquiries using various search criteria are made. The data collection procedures, the operational aims of the organization and the type of

data maintained is quite unique to the credit bureau but the basic requirements of the identification technology are not.

Many private sector organizations will use some type of unique code to identify individuals on file (a department store will have an account number for example). This code will be readily available to link individuals because it is to the benefit of the individual to supply it. A credit bureau, on the other hand, does not always rely on such a unique code to identify individuals.

Toronto Credits Limited is a private sector organization that has developed an identification technique which could be considered appropriate for use in justice information systems. Toronto Credits offers on-line credit reporting services to member bureaus such as the Ottawa Credit bureau and any credit grantor operating within the confines of the Consumer Protection Act. This service is offered for a fee to clients. An individual may also request access to his/her personal files on request.

The data concerning individuals is collected from various sources, including:

- writs;
- bankruptcy discharges;
- chattel mortgages;
- credit applications; and,
- clients.

A unique multi-level index data base using a direct access method (BDAM) has been developed by Toronto Credits to store and access data. A Consumer Record consists of mostly variable length chained segments containing data pertaining to name, personal identification, employment history, record of inquiry access, address, trade or account segments and various other segments. The Consumer Record is stored using a Control Number, which points to the record. The data

can be accessed either directly using the Control Number or through the index by supplying surname and street address and, optionally, first name, sex, spouse name, age, or Social Insurance Number. A phonetic algorithm is applied to the name and address, before the index is searched. A score is applied to each key data element match or mismatch and the records associated with the four highest scores above a pre-defined minimum score are displayed.

The Toronto Credits Consumer Record system operates in a shared service bureau environment on a large mainframe IBM model 3080A computer.

5.13 GIRLS - A Record Linkage Application

GIRLS, a Generalized Iterative Record Linkage System, is a package designed by Statistics Canada to determine likely record matches on files which have no uniquely coded record identifier. GIRLS is currently being used for statistical applications requiring file matching and, in its present form, is not suitable for use in operational systems requiring single record matching. The basic concepts of probabilistic record linkage used in GIRLS are relevant to this project however, and many of the supplied routines for pre-processing of files can be applied to any information system that requires field preparation prior to matching, be it record search matches or data file record linkages.

The GIRLS system includes several processing phases. The following describes, in general terms, the functions of each phase:

1) Pre-Process Phase — This step is not currently an integral part of the GIRLS system but routines are available to the user to prepare the file(s) which may include standardization of name and address fields and data conversion.

2) Analysis Phase — Program source code and tables are generated based on user supplied file definitions and field comparison rules.

3) Weight Generation Phase — Based on frequency counts, weights are generated and stored for the various data fields involved in the linkage.

4) Compare Phase — Potential links of record pairs are determined based on user-defined field by field comparison criteria. The records are grouped into 'pockets' or blocks of records depending on values in certain key fields and are compared physically only within these 'pockets'. Information concerning potential links is stored on the GIRLS project data base.

5) Weight Phase — This phase refines the total weight for the linked pair.

6) Link Phase — Based on the total weight determined in the previous phase and user-defined threshold weights, the system determines whether the pair is a rejected, possible or definite link.

7) Group Phase — The system identifies sets of records which are tied together through definite and/or possible links. 'Conflicts' can be automatically resolved by the system or manually resolved by the user. The information concerning group relationships is stored on the project database.

8) Report Phase — Information is produced at a detail level concerning links, weights and groups.

9) Info Phase — Information is provided on the current status of the linkage run.

10) Post-process Phase — Not a component of GIRLS, this phase is the final operation of linkage. Using the information on the GIRLS reports and the relationships on the project database, the user is able to perform functions such as building composite records or creating summary files for data analysis.

GIRLS allows the user to interactively adjust parameters until optimum linkage results are achieved. Unwanted repetition of successfully completed steps is unnecessary. The iterative process is usually run in on-line mode against a sample of records. Adjustments to parameters are set at the terminal to allow the user to experiment with the record sample until the optimum linkage results are obtained. At this point, the user requests a full run in batch mode to obtain printed reports for reference when determining the post-processing steps to be taken.

A number of routines have been developed for use in the pre-processing phase of a GIRLS application. The routines are written as procedures to be run with a user-supplied driver program and are available to the user to perform specific field standardization and conversion processes. Since these routines are not specific to a particular application, they are appropriate for use in any information system that must identify individuals. Comprehensive routines for condensing and standardizing both addresses (ASKGEN2, ENCODA) and names (NSKGEN5) have been developed to perform pre-processing for the Business Register Maintenance System, an application using the GIRLS package. The address routine, in particular, is worthy of mention in that it generates a very reliable search key that minimizes both the format and spelling errors that commonly hinder searches and matching using an address as a key.

5.14 SUMMARY TABLE OF CHARACTERISTICS

	Phonetic Routine	Optional Identifiers	Scoring	Individual Based	Documentation Available for Further Study
British Columbia Provincial Case File System	Internally Developed	Yes	No	Yes	Yes
Alberta Court Automated Project	FIND-Based	Yes	No	No	Yes
Alberta Corrections Management Information System	Russell Soundex	Yes	No	Yes	Yes
Saskatchewan Corrections	None	Yes	No	Yes	Yes
Ontario Corrections	Internally Developed	Yes	No	Yes	Yes
New Brunswick Justice Information System	NYSIIS-Based	No	No	No	Yes
Nova Scotia Social Services	Internally Developed	Yes	No	Yes	Yes
RCMP Persons	FIND	Yes	Yes	Yes	Yes
Utah Juvenile Justice System	NYSIIS	Yes	Yes	Yes	Yes
Maricopa County Juvenile On-line Tracking System	Russell Soundex	Yes	No	Yes	Yes
Memphis Juvenile Justice System	Internally Developed (JISRA)	No	No	Yes	No*
Toronto Credits Limited	Internally Developed	Yes	Yes	Yes	No

Girls - N/A

* Documentation is available if the JISRA package is purchased.

6. TECHNOLOGY DETAILS

6.0 Introduction

This chapter provides a more detailed description of the identification techniques employed in five of the systems reviewed in Chapter 5, Technology Summaries.

The purpose of this study is to provide a review of existing identification technology upon which to base subsequent phases of the study. To this end, the systems included in this section are not necessarily the 'best' integrated information systems available; rather, they represent a mix of the different approaches implemented in the area of identification technology.

6.1 Ontario Corrections

The Rapid Enquiry System (RES) provides on-line enquiry access to basic data concerning current and past inmates of the adult correctional system. Inquiries to RES are made to establish prior contact and generate requests for complete client profile data from the batch Adult Information System.

The RES online data base is accessed by a clerical Inmate Records Group whenever an inmate enters a jail to establish whether that inmate has been in the system before. Searches can be performed using one of four types of keys:

1) name/sex/birth date or age;
2) Main File Number (the primary key of both RES and AIS);
3) institution or probation and parole number; and,
4) Fingerprint Serial Number.

The name/sex/birthdate key can be entered as full or partial keys as follows:

Last Name/First Name/Birth Date;
Last Name/First Name;
Last Name/First Name/Sex;
Last Name/First Name/Sex/Birth Year;
Last Name/First Name/Sex/Birth Month/Birth Year; and,
Last Name/First Name/Age (converted to birth year ± 3).

A phonetic routine is applied to the first and last name to derive a key which is then compared to the equivalent key of the data base. The operator is required to enter the full first and last names to ensure the integrity of the phonetic key.

The basic phonetic key transforms all letters following the first letter as follows:

A E H I O U W Y	TO	SPACES
C G J K Q X	TO	1
L	TO	2
M	TO	3
N	TO	4
B P	TO	5
R	TO	6
S Z	TO	7
D T	TO	8
F V	TO	9

This transformation gives the coded value of the letter under normal circumstances. There are times when the letter, in the context of surrounding letters, takes on another sound or no sound at all. A comprehensive set of exception routines is in place to convert letters under special circumstances. This will ensure that commonly mispelled names such as "JERRY" and "GERRY" will be matched.

The resulting key after the transform and conversion routines will be in the form of:

SURNAME CODE	A9999
FIRST NAME CODE	A9
2ND GIVEN NAME	A9 (may be spaces or zeros)

Partial searches can be performed using name only; name and sex; name, sex, birth year; or name, sex, birth year, birth month. If the age is entered rather than the birth date, a variation of up to three years is allowed on a match.

The RES on-line data base is not actually updated at the terminal. The AIS batch master file will be updated with the new data and will pass the inmate information back to RES during the next update.

6.2 Nova Scotia Social Services

The 'Client Index System' in Nova Scotia uses a straightforward method for determining prior contact which is suited to the requirements of an integrated information system serving a smaller-scale client population.

A prior contact search can be initiated by supplying either the unique client index number if known or personal attribute identifiers. As an option the Social Insurance Number can also be entered. The remainder of this section deals with the manner in which a match is attempted using personal attribute identifiers.

The client identifying information required by the system is surname and, optionally, given name or initials, sex, age or date of birth.

The soundex key is made up of the first initial of the surname and the consonants are numerically represented. The vowels are dropped. Also contained in the key is a sex code and the year of birth. The file is

searched using this key, and any other identifiers supplied are checked.

SURNAME - If the surname is entered, the system will list all clients having that surname. If the surname is entered along with the given name, all clients matching will be listed. The surname can be entered along with the given name initial; all clients with the same first initial and surname will be listed. Finally, up to four given name initials, separated by commas, can be entered in conjunction with the same surname. All clients having different given name initials but the same surname will be listed.

GIVEN NAME - If the given name is supplied an exact match, only, is considered a hit on this element. Partial name searches are not performed.

SEX - "M" or "F" must be entered. "U" (Unknown) can be entered to perform prior contact checks but cannot be recorded on a record as a prime identifier.

AGE OR BIRTHDATE - The system will search over three calender years - one year prior to the birth year provided (or calculated if the age is entered) and one year following the birth year. The specific year and the month can also be entered.

In the event of only one matching file record, a client profile is displayed for that individual. If more than one possible match is found, a "short list" of individuals is displayed. The operator will then select individuals from the "short list" for a more complete display of client profile data.

The client profile data which is maintained by the Client Index System points to the individual's program involvement. Each agency maintains detailed files concerning an individual's activity within the program.

6.3 Royal Canadian Mounted Police (RCMP)

The on-line file searching technique used in the RCMP PERSONS system is one of the most reliable namesearching systems currently in use. The phonetic routine, commonly referred to as the FIND code (Fonetically Indexed Name Directory) has been adopted for use in numerous systems including the Provinces' Registered Owners systems, the Edmonton City Police and National Archives.

The RCMP namesearching system is comprised of four basic steps.

1) Determine the Prime Code:

A phonetic routine is applied to the surname to determine the Prime code. This code points to the segment of the file to be searched.

2) Determine the XREF (cross-reference) Find Codes:

Alternate phonetics are applied to the surname and the possible combinations of the alternate (XREF) code and the Prime FIND code point to alternate segments of the file to be searched.

3) Scoring Search:

The file segments are searched, scores are assigned to matching, mismatching or missing key data elements and the total record score is computed.

4) Listing:

The probable matches are returned in descending order of score.

The purpose of name coding (Step 1 and Step 2) is to segment a name file and to place like-sounding names in the same segment. Segmenting, or Name Grouping as it is sometimes referred to, reduces the number of physical record searches to determine a match. The FIND code phonetically translates the surname to a five character code which points to the segment of like-sounding surnames. The first letter of the surname is translated to one of the characters:

B, D, E, F, J, K, L, M
N, P, R, S, T, U, W

and the remaining letters are translated to digits.

The problem with any phonetic routine is that there comes a point where further refinements to the code to handle known name variations will degrade the coding of existing variations. To avoid this, an XREF code (cross-reference) is generated to handle the variations. The characters in the Prime code and the XREF code are then cross-referenced to produce up to 64 combinations of codes pointing to the segments of the file to be searched.

Once the segments have been located, the record scores are generated by applying a probability scoring technique to each key data element. Scores are calculated for matches, mismatches or missing elements using the following identifers:

SURNAME*
TEETH-TREATED
TEETH-ABSENT
DATE OF BIRTH*
AGE
GIVEN NAMES*
PLACE OF BIRTH*

EYES
HEIGHT
WEIGHT
RACE
SEX

The * denotes elements that gain points on a "hit" or match. The others lose points if they do not match within a range.

The probability scoring method used for the surname is explained more fully to provide some insight into the manner in which some scores are assigned. It should be remembered that each data element score is calculated so that any single mismatch will not exclude the file record (the exception to this is sex, where a mismatch excludes the record from further consideration).

Every record in the segment to be searched will automatically have a phonetically similar name to the subject individual's surname. An exact match within the segment on an uncommon surname will be assigned a higher score than an exact match on a common name. To do this, the segment size of the largest segment is compared to the number of occurrences of the subject individual surname. This ratio is then assigned a score based on a scale developed by Dr. Howard Newcombe called the Newcombe Scale.

Example:

RATIO	SCORE
1-1	0
1-2	1
1-4	2
1-8	3
...	...
1-8192	13

Largest Segment Size 14,737 entries

Surname - Smith 11,449 entries
Ratio 14,737 : 11,449 or 1 to 1
Score = 0

Once the scores are assigned to each data element, the total record score is computed. The probable matches are then returned in decreasing order of probability sequence. To aid in judgement, the record score achieved as well as the highest record score possible for the particular set of identifiers are returned.

6.4 Utah Juvenile Justice System

The Utah Juvenile Justice System uses the namesearching routines of a stand-alone identification processing system called MICRONYM. MICRONYM, developed by SEARCH Group Inc. of Sacramento, California, is a completely stand-alone system for the management and operation of document processing and namesearch within a state identification bureau.

Utah has modified the MICRONYM software for use on its mainframe but the concept of a separate master name index remains unchanged. A unique case number is the key which links subject records of the same individual together. The case number is automatically incremented by one for each new subject individual.

The first step of the namesearch function is to determine the Name Group of the records to be searched. The Name Grouping method used in MICRONYM was developed in 1970 for the New York State Information and Intelligence System (NYSIIS) and is commonly known as the NYSIIS Name Grouping method. The process is described in the following steps:

1. If the first letters of the name are:
   "MAC" then change these letters to "MCC"
   "KN" then change these letters to "NN"
   "K" then change this letter to "C"
   "PH" then change these letters to "FF"

"PF" then change these letters to "FF"
"SCH" then change these letters to "SSS".

2. If the last letters of the name are:
"EE" then change these letters to "Y"
"IE" then change these letters to "Y"
"DT" or "RT" or "RD" or "NT" or "ND" then change these letters to "D"

3. The first character of the phonetic code is the first character of the input name.

4. In the following rules, a scan is performed on the characters of the name. This is described in terms of a program loop. A pointer is used to access to the current position under consideration in the name. Step 4 is to set this pointer to point to the second character of the name.

5. Considering the position of the pointer, only one of the following statements can be executed.

If the last character of the name has been scanned then go to rule 7.

If the current position is a vowel (AEIOU):
then if equal to "EV" then change to "AF"
otherwise change current position to "A".

If the current position is the letter:
"Q" then change the letter to "G"
"Z" then change the letter to "S"
"M" then change the letter to "N".

If the current position is the letter "K":

then if the next letter is "N" then replace the current position by "N" otherwise replace the current position by "C".

If the current position points to the letter string
"SCH" then replace the string with "SSS"
"PH" then replace the string with "FF".

If the current position is the letter "H" and either the preceding or following letter is not a vowel (AEIOU) then replace the current position with the preceding letter.

If the current position is the letter "W" and the preceding letter is a vowel then place the current position with the preceding position.

If the current position is the character space or "-" then replace the current position with the preceding position.
If none of these rules applies, then retain the current position letter value.

6. If the current position letter is equal to the last letter placed in the code then set the pointer to point to the next letter and go to step 5.

The next character of the code is the current position letter.

Set the pointer up by one to point at the next letter.

Go to step 5.

7. If the last character of the code is the letter "S" then remove it. Go to step 8.

8. If the last two characters of the code are the letters "AY" then replace them with the single character "Y". Go to step 9.

9. If the last character of the code is the letter "A" then remove this letter.

The resulting code is a string of characters of variable length. It is possible to create a null code in which no letters of the name were retained in any form. The null code could be created by a name containing only vowels and a terminal "S". If a blank code is generated, the first three letters of the name are used as the code.

10. Place the first initial in code position 10 of 10 character code, and truncate if necessary to 10 characters.

Once the Name Group has been determined, each record is searched and the scores are assigned to the following data elements:

Surname
First Name
First Initial
Middle Name
Middle Initial
Sex
Race
Month of Birth
Day of Birth
Year of birth
Fingerprint Pattern
Social Security Number

In MICRONYM, there is one match score and one mismatch score for each data element. There is no determination of "severity of wrongness". A missing element will score zero. A mathematical model called the Bayes Theorem was chosen to determine the quantitative scores for each element and the aggregate record scores. If a pre-set threshold score is not met, the file record is considered a "no hit".

The results of the namesearch are returned to the operator, who will then inform the system of the results of search verification. A new master subject record may be created, alias records may be created or blank portions of an existing record may be replaced with new information.

6.5 Maricopa County Juvenile On-line Tracking System (JOLTS)

The identification files of the JOLTS database network are organized by a phonetic system which is very similar to the original Russell Soundex code. Each juvenile has a record containing detail identifying data which acts as a cross-reference to the file numbers assigned to their names. These detail records are grouped together under "Soundex-Master" records according to the sound of their names, rather than the spelling. There are approximately 3,200 Soundex-Master records and 45,574 Soundex-Detail records on the JOLTS data base.

The date elements used to initiate a search are:

Last Name (Mandatory)
First Name (Optional)
Sex (Optional)
Date of Birth (Optional)

The Soundex system converts the Last Name to a 4-character code using the following criteria:

1) First character is first character of Last Name

2) Second through Fourth characters are zero-filled then computed as follows:

a) A, E, I, O, U, Y, H, W are bypassed.
b) B, P, F, V converted to 1.

c) C, G, J, L, G, S, X, Z converted to 2.
d) D, T converted to 3.
e) L converted to 4.
f) M, N converted to 5.
g) R converted to 6.

This code is then used to directly retrieve a Soundex-Master record. Soundex-Detail records are then retrieved sequentially and displayed on the screen in the following order:

1) Exact matches on Last Name, First Name, Sex, and Date of Birth.

2) Exact matches on Last Name, First Name and Sex.

3) Exact matches on Last Name, First Initial and Sex.

4) Exact matches on first four characters of Last Name and Sex.

5) All other records that are detail to the Soundex-Master.

The output display of the search results include the following data fields:

Last Name
First Name
Sex
Date of Birth
Race
Case Type
Alias
Mother's First Name
File Number

Quality control reports are run monthly to report possible duplicate situations (the same juvenile with more than one file number).

APPENDIX A - LIST OF CONTACTS

D. Anderson	Saskatchewan Department of Social Services
O. Anderson	Alberta Solicitor General
J. Barrett	Maricopa County Juvenile Court, Phoenix, Arizona
R. Beattie	Canadian Centre for Justice Statistics
L. Bennett	Royal Canadian Mounted Police
D. Brown	Canadian Centre for Justice Statistics
S. Burbridge	Solicitor General Canada
J. Burns	New Brunswick Department of Justice
R. Carter	Alberta Attorney General
W. Cook	Toronto Credits Limited
D. Clark	New Brunswick Department of Justice
T. Conrad	Nova Scotia Department of Social Services
F. D'Argis	British Columbia Ministry of Attorney General
A. Desmarais	Statistics Canada
B. Doherty	Juvenile Courts, Memphis, Tennessee, USA, Juvenile Information Systems & Records Access (JISRA)

APPENDIX A (Cont'd)

R. Ferguson	Saskatchewan Department of Justice
J. Fizzera	Nova Scotia Department of Social Services
R. Girard	Ministère de la justice, Montréal, Québec
B. Grainger	Canadian Centre for Justice Statistics
T. Hill	Statistics Canada
D. Jacobson	Saskatchewan Department of Justice
F. Kaptein	Systems West Limited, Edmonton, Alberta
I. Kowaliw	Toronto Credits Limited
N. Lee	Canadian Centre for Justice Statistics
F. Letkeman	British Columbia Ministry of Attorney General
J. Linton	Ontario Ministry of Community and Social Services
J. Maginski	Reader's Digest, Montreal, Québec
D. Martin	Manitoba Corrections and Community Services
C. McCartney	Ontario Ministry of Community and Social Services
Inspector J. McDougall	Royal Canadian Mounted Police

APPENDIX A (Cont'd)

Mr. Morin	Ottawa Credit Bureau Limited
B. Muhly	Alberta Attorney General
J. O'Gorman	Ontario Ministry of Correctional Services
R. Ouellet	Ministère de la Justice du Québec
M. Phillips	Utah Juvenile Courts, Salt Lake City, Utah, USA
J. Pirozzi	National Council of Juvenile and Family Court Judges, Reno, Nevada, USA
L. Radulescu	Ottawa Police Department
E. Rathbun	National Crime Information Centre, New York, USA
J.A. Schriel	Health and Welfare Canada
L. Shapiro	Ontario Ministry of Community and Social Services
S. de Silva	Canadian Centre for Justice Statistics
Deputy Chief S. Stevens	Edmonton Police Department
T. Stickland	Saskatchewan Department of Social Services
J. Townesend	Solicitor General Canada
P. Trommelen	Alberta Department of Social Services

R. Warner	Manitoba Corrections and Community Services
T. Wilson	SEARCH Group, Inc., Sacramento California, USA

Statistics Canada
Canadian Centre for Justice Statistics

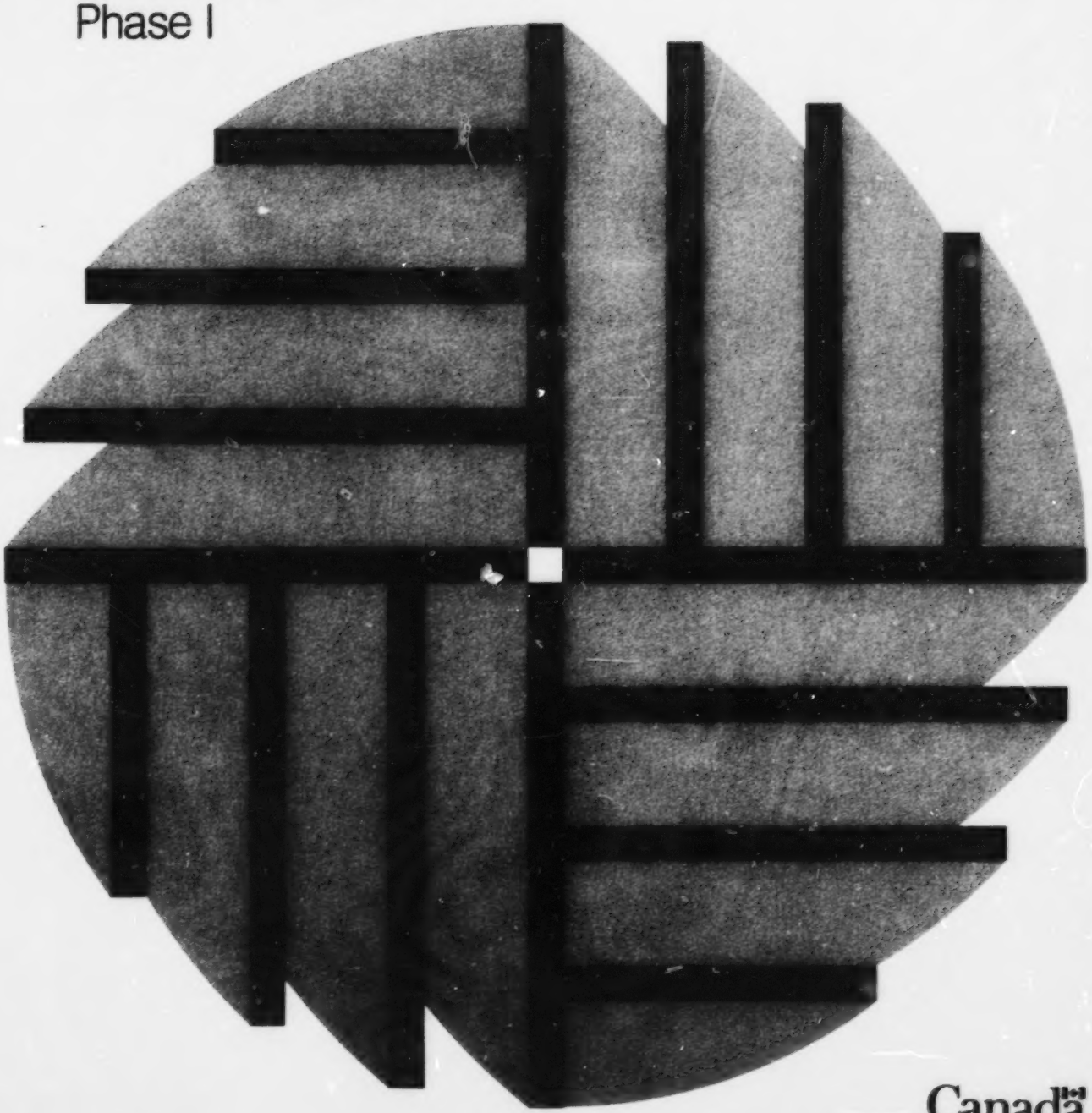
Statistique Canada
Centre canadien de la statistique juridique

R-051B
(Mars 1985)

c.2

Étude de la technologie d'identification indépendante des empreintes digitales

Phase I



Canada

ETUDE DE LA TECHNOLOGIE D'IDENTIFICATION INDEPENDANTE DES EMPREINTES DIGITALES

PHASE 1

REALISEE POUR LE

CENTRE CANADIEN DE LA STATISTIQUE JURIDIQUE

ET

FINANCEE PAR LE SOLLICITEUR GENERAL DU CANADA
DANS LE CADRE DU PROGRAMME DE CONTRIBUTION A L'ETABLISSEMENT
DE SYSTEMES POUR L'APPLICATION DE LA LOI SUR LES JEUNES CONTREVENANTS

PAR

SYNERLOGIC INC.
(ANCIENNEMENT BAILEY & ROSE LTD.)

1er MARS 1985

ÉTUDE DE LA TECHNOLOGIE D'IDENTIFICATION INDÉPENDANTE DES EMPREINTES DIGITALES
PHASE 1

TABLE DES MATIERES

		Page
1.	**Précis**	1
2.	**Introduction**	3
2.1	Historique	3
2.2	Objectifs et activités	3
3.	**Caractéristiques des technologies**	4
4.	**L'identification**	7
5.	**Exposé sommaire des technologies**	11
5.0	Introduction	11
5.1	Le Provincial Case File System de Colombie-Britannique	12
5.2	Le Court Automated Project d'Alberta	14
5.3	Le Corrections Management Information System d'Alberta	15
5.4	Services correctionnels de la Saskatchewan	17
5.5	Services correctionnels de l'Ontario	18
5.6	Système d'information juridique du Nouveau-Brunswick	20
5.7	Services sociaux de la Nouvelle-Écosse	22
5.8	Gendarmerie royale du Canada	23
5.8.1	Le fichier des personnes de la Gendarmerie royale du Canada (GRC)	23
5.8.2	Autres technologies de la Gendarmerie royale du Canada (GRC)	25
5.9	Système judiciaire pour les jeunes de l'Utah	26
5.10	Juvenile On-Line Tracking System du comté de Maricopa	28
5.11	Système judiciaire pour les jeunes de Memphis - Application de JISRA	30
5.12	Toronto Credits Limited - Application du secteur privé	32
5.13	GIRLS - Application de liaison des dossiers	34
5.14	Sommaire des caractéristiques	39
6.	**Détails des technologies - La liste abrégée**	40
6.0	Introduction	40
6.1	Services correctionnels de l'Ontario	40
6.2	Services sociaux de la Nouvelle-Écosse	42
6.3	Gendarmerie royale du Canada	44
6.4	Système judiciaire pour les jeunes de l'Utah	48
6.5	Juvenile On-line Tracking System du comté de Maricopa	52
	Annexe A - Liste des contacts	55

1. PRÉCIS

Le présent rapport donne les résultats de la Phase 1 de "l'Étude de la technologie d'identification indépendante des empreintes digitales", projet du Service d'assistance technique du Centre canadien de la statistique juridique. Le projet a été mis en route parce qu'un certain nombre de juridictions s'intéressaient à élaborer des systèmes d'information intégrés pour faciliter l'application de la Loi sur les jeunes contrevenants (LJC). Lorsque le sujet est un individu, un système intégré d'information doit pouvoir identifier cet individu à l'intérieur du système. Les empreintes digitales constituent un identificateur fiable et positif, mais ne peuvent servir que pour une petite proportion des contrevenants.

Les autres méthodes présentement en usage pour identifier les individus ont été examinées lors de cette phase du projet. L'examen a révélé que seule la recherche onomastique (l'emploi du nom de l'individu et d'identificateurs secondaires facultatifs comme la date de naissance et le sexe) sert de solution de rechange à l'identification par les empreintes digitales.

Le présent rapport donne une vue d'ensemble technique des diverses applications de la technologie de la recherche onomastique. Le chapitre 4, qui traite de l'identification, pourra intéresser les lecteurs qui désirent se familiariser avec les concepts de base de la recherche onomastique et le dossier de l'identification. Le chapitre 5 donne une vue d'ensemble de treize systèmes d'information qui utilisent la technologie de recherche onomastique tandis que le chapitre 6 donne des descriptions techniques plus détaillées des techniques particulières de cinq systèmes.

Avant de paraître, le présent rapport a été distribué dans l'ensemble du Canada pour permettre de le commenter et de l'examiner. Il présente deux options quant à l'avenir de la recherche sur la technologie de la recherche onomastique:

1) Les provinces et territoires peuvent utiliser le présent rapport à titre d'information et étudier indépendamment la valeur des technologies par rapport à leurs propres travaux d'élaboration de systèmes. Aucune autre mesure ne sera prise par le Centre canadien de la statistique juridique.

2) Le Centre canadien de la statistique juridique poursuivra son étude des cinq technologies traitées au chapitre 6. Ce travail aboutira à un document technique susceptible de servir à l'élaboration détaillée des systèmes.

Les secteurs de compétence ont fait savoir qu'il était souhaitable que le Centre canadien de la statistique juridique entreprenne une étude au cours de la phase 2. Le deuxième rapport renfermera les renseignements suivants:

- les interfaces entre la fonction d'identification et l'ensemble des éléments fonctionnels du système d'information;

- le détail de la conception de la fonction d'identification;

- la transférabilité ou la flexibilité du code de programme ou du langage de programmation;

- les problèmes que posent les technologies, y compris les facteurs de risque et les taux de réussite déterminés par les experts qui s'occupent de ces technologies; et

- les poids et contrepoids mis en place pour réduire les risques reconnus.

2. INTRODUCTION

2.1 Historique

La proclamation de la LJC a ravivé dans les provinces l'intérêt envers les systèmes intégrés d'information juridique. Les dispositions de la LJC portant sur la conservation et la destruction des dossiers, en particulier, ont fait ressortir la nécessité d'intégrer les renseignements concernant un individu quel que soit le secteur de l'appareil judiciaire en cause et quel que soit le moment ou le lieu de l'événement.

Un élément clé d'un système intégré d'information, lorsque le sujet est un individu, est l'identification de cet individu. Dans un système axé sur un secteur et portant sur un événement isolé, l'identification de l'individu ne pose aucun problème important, car il y a progression naturelle de l'individu à travers le système. Toutefois, un bon système intégré doit pouvoir identifier précisément un individu, quel que soit le secteur, le moment ou le lieu. La mise en place d'un bon système d'identification, assorti de diverses mesures de contrôle, exige beaucoup de temps et de soin.

Plusieurs organisations du Canada et des Etats-Unis ont mis en place des systèmes d'information où l'identification de l'individu se réalise sans l'aide d'empreintes digitales. Le monde judiciaire peut profiter de l'expérience des personnes chargées de ces systèmes pour élaborer des systèmes d'information répondant aux exigences de la LJC.

2.2 Objectifs et activités

L'objectif du projet, selon la proposition de projet, est le suivant:

"Apprendre à partir de l'expérience des autres, les avantages et les désavantages de la technologie indépendante des empreintes digitales et en quantifier les risques".

Le présent rapport donne un aperçu des technologies actuelles d'identification qui sont indépendantes des empreintes digitales et qui pourraient servir à des systèmes d'information portant sur les jeunes.

La Phase 1 de l'étude a porté sur les sujets suivants:

- Quels systèmes d'identification sont en place et où?
- Quelles sont les caractéristiques de ces systèmes?
- Quel est le rapport fonctionnel entre l'élément d'identification et les fonctions globales du système d'information?

Le projet comportait les activités suivantes:

- entrevues avec des employés du Centre canadien de la statistique juridique;
- entrevues (surtout par téléphone) avec des personnes au courant des systèmes d'identification du Canada et des Etats-Unis;
- évaluation de l'état actuel du dossier de l'identification;
- recherche de documentation sur les treize systèmes existants; et
- évaluation et nouvelle documentation d'un groupe représentatif de cinq systèmes.

3. CARACTÉRISTIQUES DES TECHNOLOGIES

La présente phase de l'étude n'évalue pas les systèmes d'information et les technologies d'identification qu'ils comportent expressément en fonction des besoins des systèmes provinciaux d'information. Ce sont, le cas échéant, les phases subséquentes du projet qui traiteront de l'adaptabilité aux systèmes provinciaux ou territoriaux, des facteurs de risque et des taux de réussite. Une autre étude également entreprise par le Centre canadien de la statistique juridique, soit l'évaluation préliminaire des progiciels ayant trait à la Loi sur les jeunes contrevenants, devrait compléter cette analyse des techniques d'identification. La présence phase examine les technologies actuelles d'identification dans le cadre de l'ensemble du système et fait rapport d'un échantillon représentatif de ces systèmes, comme point de départ d'autres études.

Pour accomplir cette tâche, il est nécessaire de dégager une liste générale des besoins pour assurer que les systèmes étudiés au cours de cette phase fourniront effectivement un point de départ pour des études ultérieures. La liste suivante a servi de guide général permettant de juger s'il convenait d'inclure un système donné dans le présent rapport:

- Les technologies doivent pouvoir emmagasiner et retrouver des renseignements sur des individus par des moyens autres que les empreintes digitales.

- Les technologies doivent être capables de dépistage. L'identification de l'individu doit être positive quel que soit le moment ou le lieu de l'événement. On ne peut postuler le dépistage central de la cueillette et de la coordination des données.

- La technologie d'identification doit, de façon générale, convenir aux systèmes d'information portant sur les jeunes. Il se peut

que des technologies fondées dans une large mesure sur des caractéristiques qu'on ne retrouve pas d'ordinaire chez les jeunes ne soient pas appropriées. Par exemple, des technologies qui utilisent le numéro d'assurance sociale ou le numéro du permis de conduire pourraient ne pas convenir à des systèmes d'identification des jeunes.

- Le système d'informaton devrait être conçu à des fins opérationnelles aussi bien que statistiques. Un système purement statistique peut avoir une tolérance plus grande du risque au titre du dépistage et de l'intégrité des données.

Le mandat de l'étude est d'assurer une large représentation des technologies actuelles d'identification qui puisse servir de point de départ à des analyses futures. Divers systèmes du Canada et des Etats-Unis ont fait l'objet d'examen et de documentation et une "liste abrégée" de cinq systèmes a été choisie pour une étude plus poussée. Elle vise à donner la meilleure représentation possible des technologies. Elle a été établie d'après les critères suivants:

- Le système judiciaire comprend les tribunaux, les services correctionnels et l'application de la loi. Il devrait y avoir une technologie représentative pour chaque secteur.

- La recherche préliminaire sur la quesiton de l'identification et sur les systèmes existants a montré que diverses approches ont été adoptées (voir le chapitre suivant). Des applications de ces diverses approches devraient faire l'objet d'une étude ultérieure.

- La "liste abrégée" peut comprendre des systèmes en place soit au Canada soit aux Etats-Unis. Bien que les systèmes américains

figurant dans l'étude puissent être très bien développés sur le plan du système global, il est devenu manifeste que l'élément d'identification de certains systèmes canadiens peut soutenir la comparaison, sur le plan technologique, avec les systèmes américains.

4. L'IDENTIFICATION

Une identification précise est essentielle à la réussite de tout système d'information axé sur l'individu. Dans le domaine des casiers judiciaires, la fonction d'identification doit pouvoir intégrer l'information concernant un individu quel que soit le secteur desservi ou le moment ou le lieu de l'événement. On a consacré une somme considérable de recherche à l'élaboration de techniques automatisées et semi-automatisées pour l'identification des auteurs d'infractions graves au moyen des empreintes digitales. La classification fondée sur les empreintes digitales ne constitue un moyen fiable et positif d'identification que pour une petite partie des contrevenants. Les infractions moins graves doivent toujours être reliées à l'individu. On y arrive au moyen d'une technique appelée la recherche onomastique, au moyen du nom de l'individu et d'identificateurs des caractéristiques personnelles comme le sexe ou la date de naissance.

Les systèmes de recherche onomastique fondés sur les caractéristiques personnelles n'utilisent pas d'ordinaire ces caractéristiques personnelles comme élément identificateur unique et primaire fondant les liaisons entre les dossiers au sein des fichiers informatisés. Des identificateurs codés, souvent des numéros arbitraires, peuvent être la seule clé utilisée pour emmagasiner les dossiers et, dans certains cas, pour retrouver l'information concernant un individu. L'identificateur codé, s'il est connu de l'opérateur, peut être utilisé en conjonction avec les identificateurs de caractéristiques personnelles pour effectuer des

recherches. Les identificateurs de caractéristiques personnelles servent surtout à

- établir un contact antérieur avec le système (est-ce que le système informatique possède déjà des renseignements concernant l'individu) en donnant la liste des individus figurant au fichier qui ont des caractéristiques conformes à celles du sujet. A partir de ces listes, l'opérateur utilise son jugement pour décider si oui ou non des renseignements sur le sujet ont préalablement été entrés dans le système et

- assurer un fondement pour l'organisation automatisée et efficace de la mémoire et pour la recherche.

Un rapport préparé pour le Comité Fédéral-provincial de l'information juridique en avril 1976[1] examinait la question des identificateurs et passait en revue les options suivantes:

- numéro administratif local - numéro assigné au palier local;
- numéro de série des empreintes digitales, disponible seulement pour les auteurs d'infractions graves;
- numéros assignés par le gouvernment - cette catégorie comprend le numéro d'assurance sociale, le numéro de permis de conduire, et
- nom et statistiques personnelles.

Selon l'étude de 1976, l'option la plus appropriée comme premier identificateur était le nom de l'individu. On reconnaissait que le nom ne peut que ramener l'identification à un choix. On postulait également que l'individu communiquerait volontiers ce renseignement.

---

[1] The Needs for Data Linkages in the Corrections Component of the Criminal Justice System, rapport préparé pour le Comité fédéral-provincial de l'information juridique par le groupe de travail sur les identificateurs, avril 1976.

En mars 1975, Search Group Inc. de Sacramento (Californie) publiait un rapport donnant une vue d'ensemble de l'état des connaissances en matière d'identification pour l'appareil judiciaire. Le rapport portait sur des systèmes automatisés de recherche onomastique et sur des techniques automatisées ou semi-automatisées de recherche et de prise des empreintes digitales. Le rapport déclare que "toute la recherche nécessaire dans le domaine de la recherche onomastique a été réalisée"[2]

L'examen des systèmes d'information que nous avons entrepris a révélé que les systèmes actuels d'information dans le domaine de la justice pénale continuent à s'en remettre à des techniques fondées soit sur les empreintes digitales soit sur la recherche onomastique pour identifier les individus.

Une étude plus récente, entreprise en 1982 par le Office of Technology Assessment des ETats-Unis a évalué les options pour un système pénal informatisé national. Passant en revue l'état actuel des systèmes de casiers judiciaires, le rapport déclare: "On croit à l'heure actuelle dans le monde de la justice pénale que les empreintes digitales sont le seul fondement fiable et uniforme pour une identification positive. L'échange de dossiers fondé sur les seuls noms aboutit à un pourcentage élevé d'erreurs par suite de l'usage fréquent de faux noms et des ressemblances entre bon nombre de noms de famille communs".[3] Des efforts considérables ont été consacrés à l'élaboration de techniques visant à minimiser les risques que comporte la graphie des noms de famille et les resemmblances entre ces noms. Ces techniques sont désignées sous le nom de recherche onomastique dans le reste du rapport. Les systèmes d'information étudiés ici utilisent tous une technique quelconque de recherche onomastique.

---

[2] Search Group, Inc., Search Group Masterplan for Identification Systems Upgrade, mars 1975.

[3] An assessment of Alternatives for a National Computerized Criminal History System, projet du Office of Technology Assessment, Congrès des Etats-Unis, octobre 1982.

Le reste du chapitre vise à mettre le lecteur au courant des approches fondamentales utilisées en recherche onomastique et des termes servant à décrire les méthodologies employées par les systèmes. Il ne faut pas oublier que le but de la recherche onomastique est d'établir si le système informatique comporte des fiche sur l'individu. Toute action subséquente dépend entièrement du système.

Le terme de recherche onomastique pourrait englober les processus suivants:

1) Déterminer le groupe de fiches à explorer.

Dans les systèmes considérables, il ne serait pas souhaitable de fouiller matériellement toutes les fiches en mémoire pour procéder à des appariements. On utilise souvent une technique de regroupement des noms pour segmenter le fichier afin de réduire le nombre de fiches à dépouiller.

2) Recherche du groupe de fiches.

A cette étape, à la fois le nom et, le cas échéant, des identificateurs supplémentaires ou secondaires servent à déterminer les appariements possibles entre le sujet et les fiches. Il y a en général deux sortes de méhodes de recherche.

La recherche par exclusion met de côté toute fiche qui ne s'apparie pas à un argument de recherche. Dans ce cas, on tente d'apparier exactement, ou dans une certaine limite de tolérance, d'appariement est considéré un échec.

La recherche par cotation tient compte de tous les identificateurs lorsqu'elle tente un appariement. Chaque élément reçoit une cote, selon qu'il s'agit d'un appariement, d'un non-appariement ou d'un élément manquant, et la somme de ces cotes indique la probabilité d'appariement de la fiche.

3) Listes

Les étapes de groupement et de recherche aboutissent à une liste. En général, tous les appariements figurent à la liste s'il s'agit d'une recherche par exclusion. Dans le cas d'une recherche par cotation, deux types de listes peuvent être produits, soit une liste des fiches en ordre de leur cote ou une liste de toutes les fiches dont la cote dépasse un seuil préétabli.

4) Vérification

La décision finale d'identification est d'ordinaire une procédure manuelle.

Le regroupement des noms se fait souvent en regroupant des fiches où le nom de famille présente des ressemblances phonétiques (routine ou algorithme phonétique). Les divers algorithmes phonétiques présentement en usage sont parfois désignés sous le nom de "soundex". Le premier en date de ces algorithmes est le "Russell Soundex", système élaboré pour la recherche manuelle des index de noms. Deux algorithmes phonétiques très répandus en Amérique du Nord sont le code FIND de la GRC et le code NYSIIS du New York State Information and Intelligence System (Système d'information et de renseignements de l'Etat de New York).

5. EXPOSÉ SOMMAIRE DES TECHNOLOGIES

5.0 Introduction

La présent chapitre donne une vue d'ensemble de treize systèmes d'information qui ont comme critère fonctionnel l'identification des individus sans recours aux empreintes digitales. Divers systèmes du Canada et des Etats-Unis ont été identifiés surtout au moyen de l'inventaire des systèmes judiciaires automatisés du Canada, projet du Service d'assistance technique du Centre canadien de la statistique juridique et d'après les indications des contacts. Les treize systèmes retenus ici ne représentent nullement un inventaire complet de tous les systèmes d'information qui exécutent présentement des fonctions d'identification; ils ont été choisis afin d'assurer le fondement le plus large possible pour les études ultérieures.
Pour chacun des treize systèmes, une section décrit l'objet et l'utilisation du système ainsi que les éléments fonctionnels. Les caractéristiques de base des systèmes, par rapport à leur fonction d'identification, sont ensuite résumées en tableau au numéro 5.14.

5.1 Le Provincial Case File System (PCFS) de Colombie-Britannique

Le PCFS est un système opérationnel de dépistage qui contient des données personnelles de même que des données sur les antécédents, le tribunal et le lieu pour les clients tant actifs qu'inactifs de la Direction des services correctionnels de Colombie-Britannique.

Les fonctions opérationnelles sont:

- identification des clients (pour découvrir s'il y a eu un contact antérieur);
- renseignements sur les clients;
- antécédents des clients; and
- dépistage des clients.

Outre les fonctions opérationnelles, le système comporte une fonction d'analyse statistique pour contrôler l'utilisation des services correctionnels et permettre la planification à long terme.

Le PCFS est situé à l'administration centrale, au moins un terminal se trouvant dans chacune des cinq régions de la province. Le système conserve des données sommaires concernant tous les individus qui ont eu contact avec les services correctionnels de Colombie-Britannique au cours des dix dernières années. Des renseignements plus détaillés sur les clients sont conservés dans les établissements par le Local Institution System (LIS). Il existe une interface entre le PCFS et le LIS qui permet d'assigner le numéro des services correctionnels. Ce numéro sert d'identificateur unique reliant toutes les fiches d'un individu donné.

Les fonctions de renseignement sur les clients du PCFS permettent trois options de recherche:

1) Nom exact - Au moyen du nom de famille, du prénom et de l'année de naissance. Une variance de deux ans est permise pour l'année de naissance.

2) Nom partiel - Nom de famille tronqué, initiale du prénom et année de naissance. N'importe quel nombre de lettres peut être retranché du nom de famille.

3) Soundex - Recherche phonétique au moyen d'une base de données distincte. Au départ, cette option avait trois niveaux de sélectivité. Le niveau le plus élevé aboutissait au group le plus petit d'appariements possibles, le niveau le plus bas au groupe le plus grand d'appariements possibles. Le système a été

modifié et ne permet plus que l'accès au niveau de sélectivité le plus élevé, car les autres options aboutissaient à un trop grand nombre de réponses.

Au palier des institutions, les renseignements sur les clients se limitent à l'option du nom exact et à celle du nom partiel.

Le PCFS fonctionne sur ordinateur IBM 3081 et utilise IMS ADF version 1.3.

5.2 Le Court Automated Project (CAP) d'Alberta

Le CAP est un système à l'échelle provinciale qui dessert trois paliers de tribunaux pour adultes en Alberta. Les données proviennent de 26 emplacements de la cour provinciale, de 13 emplacements de la cour du Banc de la Reine et de deux cours d'appel. Le CAP est un système judiciaire global couvrant la plupart des fonctions des tribunaux, notamment:

- établissement du rôle
- dépistage des cas
- mandats
- contrôle financier, y compris les cautions, les restitutions et les amendes
- production de documents (sommations, mandats, etc.)
- contraventions
- antécédents des prévenus

Le système de départ n'assurait que la fonction de contraventions. La Phase 2, le système actuel, est en service au complet depuis environ un an.

Les interrogations et l'entrée des données se font en direct, mais la fonction de mise à jour se fait chaque nuit par lots. Le système comporte actuellement environ 8 000 000 de fiches.

La fonction d'interrogation sur les prévenus et leurs antécédents sert à mettre en route la recherche de toutes les causes associées à un individu donné. Les identificateurs qui peuvent servir à effectuer une recherche sont: le nom de famille, le sexe, la date de naissance, le code du tribunal ou le numéro de permis de conduire.

Un code phonétique complet, semblable à celui qu'utilise la GRC mais moins raffiné, sert pour le nom de famille; toutefois, la cotation est impossible.

Le CAP fonctionne sur ordinateur IBM 3033, situé au Centre informatique d'Edmonton, utilisant TOTAL/ENVIRON.

5.3 Le Corrections Management Information System (COMIS) d'Alberta

Le COMIS a été mis en oeuvre en 1979 pour répondre aux besoins des établissements de renvoi sous garde et de correction de l'Alberta en matière d'opérations et d'analyse statistique. Les éléments opérationnels comprennent:

Dépistage des détenus - Contrôle les mouvements des détenus et prépare les listes des tribunaux et des transfèrements.

Calcul des peines - Calcule la plupart des variations des peines.

Dépistage des clients - Fait la recherche des fichiers pour établir si l'individu a déjà reçu un numéro d'identification des contrevenants.

Un fichier de recherche contenant des données qui remontent à 1977 sert de fondement à la fonction d'analyse statistique. Les données historiques de cinq années sont disponibles en direct.

Les développements en cours comprennent:

- la conversion du logiciel de la base de données (de TOTAL/ENVIRON à ADABAS/NATURAL) pour assurer plus de souplesse;
- l'addition de renseignements sur la probation;
- interface avec le Centre d'information des services canadiens de police pour permettre l'interrogation des fichiers COMIS par la police; et
- interface avec le Court Automated Project (CAP) du Procureur général afin de permettre le transfert hebdomadaire des renseignements sur les libérations de COMIS à CAP.

Deux centres de renvoi sous garde ont des systèmes opérationnels sur mini-ordinateur. D'autres établissements sont reliés en direct à l'ordinateur central pour les recherches onomastiques et pour l'inscription des admissions et des libérations.

La fonction d'identification du COMIS est l'élément-clé de la composante de dépistage des détenus. Un des objectifs principaux du COMIS est d'accumuler des antécédents complets sous un seul numéro de fichier de détenu. Puisque le taux de récidivisme dépasse 35%, il faut une technique pour établir si le détenu est un nouvel arrivant. Au départ, la procédure d'admission de l'ordinateur principal utilisait un appariement alphabétique de base des quatre premiers

caractères du nom de famille. Cette technique comportait un certain nombre de lacunes qui ont rendu nécessaire, en 1983, une révision majeure et l'adoption d'une technique de recherche onomastique phonétique. La technique améliorée est presque identique au système Russell Soundex (voir 6.6). L'identificateur primaire est le nom de famille et les identificateurs secondaires comprennant le prénom, la date de naissance (ou l'âge), le numéro d'assurance sociale, la taille et le poids. La technique de recherche du COMIS ne permet pas la cotation des probabilités.

Le système de l'administration centrals fonctionne sur ordinateur IBM 3033 et utilise présentement TOTAL/ENVIRON. Les systèmes opérationnels des établissements utilisent des ordinateurs Eclipse 350 de Data General.

5.4 Les services correctionnels de la Saskatchewan

Les services correctionnels pour adultes en Saskatchewan sont desservis par un système de traitement par lots à l'administration centrale permettant la recherche et la statistique ainsi que par un système en direct appuyant les fonctions opérationnelles des trois plus gros centres correctionnels.

Le système de l'administration centrale, le Correctional History System, a été mis sur pied en 1976 et conserve des renseignements sur les services correctionnels et la probation pour les détenus de tous les établissements de la Saskatchewan. Les données sont reçues sur papier des systèmes manuels des établissements. Les trois plus gros systèmes, qui sont automatisés, produisent des images-cartes des données courantes sur les détenus sur des bandes hebdomadaires de mise à jour pour le système de l'administration centrale. Il n'y a pas d'interfaces automatisées entre le Correctional History System et les systèmes des établissements.

Le Correctional Centre Information System (CCIS), récemment mis au point, est un système opérationnel sur mini-ordinateur pour les trois plus grandes centres. Quatre terminaux, un dans chacun des centres et un à l'administration centrale, sont reliés à un mini-ordinateur IBM modèle 8100. Des fiches des détenus sont complètes pour les cas en activité, mais on conserve des fiches schématiques pour les données des deux années précédentes. On prévoit étendre cette période à cinq ans. Les renseignements conservés par le CCIS sont recopiés sur le système par lots de l'administration centrale. Le système procède à des appariements d'après les cinq premiers caractères du nom de famille, les initiales des prénoms ainsi que le mois, la journée ou l'année de naissance.

5.5 Services correctionnels de l'Ontario

Des systèmes automatisés tant au palier de l'administration centrale qu'à celui des établissements répondent aux besoins du ministère ontarien des Services correctionnels sur le plan des opérations et des renseignements statistiques.

Un système de traitement par lots, Adult Information System (AIS), existe depuis plus de huit ans; il a recueilli des renseignement sur environ 300 000 personnes qui ont eu des rapports avec le système provincial des corrections. Situé à l'administration centrale, le système AIS fournit une base de données pour l'analyse statistique, la gestion de la probation et de la libération conditionnelle, la vérification du calcul des peines et les profils des antécédents des clients. Les données sur les détenus et les clients sont transmises au système AIS par les établissements de même que par les agents de la probation et des libérations conditionnelles.

Les gros établissements informatisés transmettent les données directement à l'ordinateur central chaque jour. Les autres établissements, de même que les agents de la probation et des libérations conditionnelles, transmettent les données au siège social sur papier. Dans un cas comme dans l'autre, le fichier principal AIS est mis à jour par lots chaque semaine.

Le Rapid Enquiry System (RES), système en direct, a pour principal objectif de permettre de déterminer si l'individu en cause a été en contact avec le système auparavant. La base de données en direct RES est un extrait des données du système AIS. Le système RES conserve des données sommaires pour chaque individu qui a été en contact avec le système correctionnel. C'est uniquement un système d'interrogation d'identification. Les données complètes sur le profil du client ou du détenu sont une fonction du système AIS.

Le Local Institution System (LIS) est un système sur mini-ordinateur qui répond aux besoins opérationnels au palier des établissements. Ce système est présentement en service en sept endroits. Les autres établissements utilisent des systèmes manuels.

Les plans de développement à long terme pour ces trois systèmes comprennent la mise à jour technologique et l'intégration en système des fonctions assurées par les trois systèmes actuels.

La fonction d'identification, assurée par le seul système RES, utilise le nom comme identificateur primaire et le sexe ainsi que la date de naissance complète ou partielle comme identificateurs secondaires. Un algorithme phonétique élaboré sert à minimiser le risque des erreurs d'appariement due à graphie des noms de famille.

Le système AIS fonctionne sur ordinateur IBM 3033, le système RES sur ordinateur MV8000 de Data General et le système LIS sur le modèle 5140

de Data General. Les établissements locaux utilisent des lignes commutées pour transmettre des données au système AIS.

5.6 Le Système d'information juridique du Nouveau-Brunswick

Il s'agit d'un système de gestion des cas portant sur les actes criminels et les infractions punissables sur déclaration sommaire de culpabilité. Il permet de créer une fiche de cas et d'y ajouter des renseignements au fur et à mesure des procédures. Le système traite annuellement environ 50 000 cas.

La pierre angulaire du système, la gestion des cas, comprend les éléments suivants:

- présentation des renseignements de base sur l'accusé;
- dates des comparutions; et
- noms des agents de police, des témoins, des juges, des avocates et des procureurs de la Couronne.

Le système assure également les fonctions suivantes:

Contrôle des amendes	- Composante financière visant à faciliter le contrôle des paiements et des amendes en souffrance.
Contrôle des cautionnements	- Composante financière visant à faciliter le contrôle des paiements.
Contrôle de la restitution	- Composante financière permettant l'enregistrement, la perception, l'indemnisation et la réconciliation des paiements.

Rôles des tribunaux - Les rôles sont préparés jusqu'à deux semaines d'avance et servent à aviser les juges, les procureurs et les agents de police.

Contrôle des mandats - Permet l'enregistrement et le suivi des mandats.

Déclarations - Déclarations statistiques à l'appui de la fonction de gestion des cas.

Gestion des fichiers de contrôle - Une série de fichiers de contrôle servent à des fins de vérification pour identifier les juges, les procureurs, les agents de police, les témoins, les articles du code criminel, les numéros de code des réglements municipaux et les lois provinciales. Le fichier des défendeurs est le seul fichier qui contienne des données personnelles de base.

Le système est en service depuis 1982. On prévoit qu'avec le temps les procureurs s'en serviront pour le dépistage.

Le système utilise le code NYSIIS mis au point pour la police de l'Etat de New York comme algorithme phonétique de recherche onomastique. L'algorithme est appliqué tant au nom de famille qu'au prénom. Aucun identificateur secondaire n'est utilisé. Le système présente tous les cas associés au nom codé plutôt que des données identificatrices individuelles.

Le système est une version modifiée du système PROMIS (Prosecutor Management Information System), logiciel américain mis au point par INSLAW Inc. de Washington (D.C.).

Des fichiers séquentiels et indexés sont utilisés en direct sur ordinateur UNIVAC 9030.

5.7 Le Client Index System des Services sociaux de Nouvelle-Ecosse

Le ministère des Services sociaux assure un vaste éventail de services sociaux aux familles, aux enfants et aux personnes défavorisées. Le ministère doit assurer non seulement la prestation la plus avantageuse des divers programmes, mais aussi la prestation du meilleur mélange de services d'une façon inégrée et coordonnée.

Le Client Index System est la composante du système d'information de gestion qui relie tous les programmes, aidant le ministère à réaliser cet objectif. Le système assure les fonctions suivantes:

- maintenir une fiche pour tous les clients;
- assurer un mécanisme pour vérifier s'il y a eu contact antérieur;
- assurer des routines d'aménagement et de détection de la fraude;
- contenir des données communes sur les clients, données qui ne sont pas particulières à un programme;
- faciliter le partage et l'échange opérationnels de renseignements entre les programmes; et
- fournir des rapports périodiques de gestion permettant le retraçage des clients dans l'axe chronologique et d'un programme à l'autre.

L'emploi du système pour vérifier s'il y a eu contact antérieur est assez simple. Chaque fois qu'un individu entre en contact avec le Ministère, on procède à une vérification pour voir s'il est connu du système. Les identificateurs primaires sont le sexe, la date de naissance, le nom de famille, le prénom ou l'initiale. Si l'individu est inconnu, on établit dans le système une fiche contenant les renseignements personnels. Une fiche de participation au programme est enregistrée séparément.

Ce système fonctionne sur ordinateur IBM 4341, tant en direct que par lots. Il utilise un fichier VSAM.

Ce système a été mis sur pied en Nouvelle-Ecosse sur un modèle québécois et a été transféré à de nombreuses reprises. Des versions de ce système sont en usage en Saskatchewan, en Alberta, en Alaska, dans le Dakota Nord et en Idaho. L'Alaska et le Dakota Nord ont remplacé le fichier VSAM par le fichier ADABAS et ont ajointé les programmes de sécurité du revenu au système. La Saskatchewan prévoit passer à ADABAS, mais utilise présentement VSAM.

5.8 Gendarmerie royale du Canada (GRC)

5.8 1 Le fichier des personnes de la Gendarmerie royale du Canada (GRC).

Il s'agit d'un système opérationnel en direct servant à identifier des individus, comme les personnes recherchées ou disparues, les personnes en libération conditionnelle ou les personnes soumises à des restrictions. Le système assure un moyen d'identifier ces individus, indique la cause et précise pourquoi ces individus intéressent les corps policiers. Seuls les agents d'application de la loi ont directement accès au système, mais d'autres organismes autorisés (les tribunaux, mais d'autres organismes autorisés (les tribunaux, les services correctionnels) peuvent y avoir accès indirectement. Les agents d'application de la loi interrogent le système de façon routinière pour établir si un individu donné intéresse la police. Par example, lorsqu'un agent arrête un individu pour excès de vitesse il lui demande, entre autres choses, son nom. Il interroge ensuite le fichier des personnes pour établir si ce

nom (peut-être qualifié par d'autres identificateurs) correspond à un individu connu du système. La recherche pourrait par exemple indiquer que la présence d'une fiche pour un individu qui est très vraisemblablement le même que celui qui a été arrêté pour excès de vitesse. Ce renseignement permet à l'agent de décider des mesures à prendre.

Les techniques mises au point par la GRC pour identifier les individus sont utilisées largement au Canada et aux Etats-Unis. Le système d'information de gestion des services correctionnels de l'Alberta, la police municipale d'Edmonton, les Archives nationales, Emploi et Immigration, les systèmes d'immatriculation des provinces et l'Agence centrale de renseignements (CIA) des Etats-Unis utilisent tous des versions de la technique de la GRC. L'application ultime de cette approche se retrouve cependant dans le fichier des personnes.

Le système des casiers judiciaires synoptiques (CJS) de la GRC utilise également cette technique pour identifier les individus. Le CJS sert d'index, renvoyant aux fiches complètes du système des casiers judiciaires (CJ2), système qui repose sur les empreintes digitales.

Dans le fichier des personnes, douze éléments de données servent de clé lors de la recherche d'un individu. Le nom de famile est une clé obligatoire et sert à segmenter phonétiquement un groupe de fiches pour la recherche. A la différence de la plupart des techniques phonétiques de regroupement des noms, celle de la GRC permet d'appliquer d'autres codes phonétiques au moyen d'une étape supplémentaire de renvoi. Un algorithme complet de cotation est alors appliqué à chaque élément secondaire de description personnelle pour chaque fiche du segment choisi afin

d'établir la probabilité d'appariement entre le fichier des personnes et le sujet. Le système donne alors les résultats de la recherche en ordre descendant de probabilité. A ce moment, l'opérateur utilise son jugement pour décider si le sujet figure au fichier.

Le fichier utilise un système de base de données mis au point de façon interne qui porte le nom de PRISM et fonctionne sur ordinateur IBM 3033 N, avec 16 mégaoctets de mémoire. La méthode PRISM d'organisation du fichier a été conçue en fonction des caractéristiques particulières des fichiers criminels.
La technique de recherche onomastique de la GRC porte le nom de code FIND et comprend trois éléments fondamentaux:

1) le codage phonétique du nom de famille pour déterminer le segment qui doit faire l'objet de la recherche;

2) XREF codage phonétique de renvoi du nom de famille pour déterminer d'autres segments à chercher;

3) la cotation qui attribue des poids aux éléments-clés de données pour déterminer la probabilité d'appariement.

Bien que d'autres systèmes utilisent présentement des versions du code FIND, la mesure dans laquelle les techniques sont appliquées est variable.

5.8.2 Autres technologies de la Gendarmerie royale du Canada (GRC)

Aux systèmes d'identification mis au point par la GRC s'ajoute un système de renseignement, le système automatisé de renseignements sur la criminalité (SARC).

C'est le seul système que nous ayons recontré au cours de notre recherche qui ne repose pas nécessairement sur un nom comme identificateur primaire. Des recherches pondérées ou d'appariement exact utilisent jusqu'à 60 clés différentes, y compris des caractéristiques personnelles comme les marques, corporelles. Le SARC mérite d'être mentionné, mais il ne serait pas à-propos de s'y arrêter davantage ici car c'est une accumulation de données (par opposition à un système d'identification individuelle) qui exige des analystes spécialisés avec une formation poussée.

Parmi les progrès technologiques de la GRC, mentionnons une nouvelle application qui utilisera les caractéristiques faciales comme identificateurs. Des photos d'identité informatisées pourraient devenir un identificateur clé tant pour le système de renseignements sur la criminalité que pour le système des casiers judiciaires synoptiques.

5.9 Le système judiciaire pour les jeunes de l'Utah (Profile II)

Le tribunal pour les jeunes de l'Etat de l'Utah utilise des ordinateurs pour recueillir des renseignements et traiter les cas depuis 1967. Le système automatisé, qui était au départ une collection du tribunal comprenant un index des jeunes, une feuille d'inculpations et des statistiques, est devenu un réseau couvrant tout l'Etat et reliant tous les éléments de l'appareil judiciaire pour les jeunes.

Profile II est un système de dossiers opérationnels visant à aider le système judiciaire pour les jeunes à accomplir ses tâches, notament:

1) Premier traitement du cas - Le système vérifie s'il y a eu contact antérieur, saisit les données statistiques et les données de traitement et produit une feuille d'inculpations à jour;

2) Le processus d'admission - La feuille d'inculpations, qui est un sous-produit du traitement initial, aide l'agent d'admission à déterminer les mesures à prendre. Le système peut produire la requête, la sommation, le rôle du tribunal, les avis d'amendes ou le calendrier des comparutions;

3) Dépistage des ordonnances - Un sommaire des mesures du tribunal est inscrit au rôle du tribunal au cours de chaque audition et est par la suite inscrit dans chaque fiche;

4) Dépistage de la détention - Les centres de détention et les services correctionnels pour les jeunes utilisent Profile II pour inscrire l'entrée et la sortie des jeunes et dresser des listes quotidiennes de la population;

5) Caractéristiques particulières - Les principales caractéristiques particulières comprennent l'accès aux casiers judiciaires des adultes, aux dossiers des permis de conduire, aux dossiers d'immatriculation des véhicules, l'établissement de

déclarations ad hoc, l'impression de chèques et la compatabilité ainsi que l'accès au National Law Enforcement Telecommunications Network (NLETS).

L'identification des jeunes au cours du traitement initial utilise une technique mise au point par SEARCH Group (Inc.) de Sacramento (Californie), le National Consortium for Justice Information and Statistics. MICRONYM, système de traitement de l'identification, est un ensemble complet permettant la gestion et l'exploitation du traitement des documents et de la recherche onomastique dans un bureau d'identification d'Etat. MICRONYM utilise la technique de regroupement des noms mise au point en 1970 pour le New York State Information and Intelligence System (NYSIIS). Une recherche par cotation mesure la probabilité que la fiche s'apparie au sujet du document de source.

Profile II fonctionne sur deux ordinateurs AMDAHL (V6 et V8) au moyen de ADABAS/COMPLETE, progiciel de télétraitement des bases de données.

Des terminaux Courier (compatibles avec l'ordinateur IBM 3270) sont situés dans 14 tribunaux pour les jeunes, six établissements correctionnels pour les jeunes, 31 points de répartition de la police et six centres de probation et de libération conditionnelle des adultes dans l'Etat. La plupart des programmes sont écrits en COBOL. NATURAL sert à préparer des programmes ad hoc, spécialisés, pour avoir accès aux données.

5.10 Le Juvenile On-Line Tracking System (JOLTS) du comté de Maricopa

Il s'agit d'un système automatisé de données pour le tribunal pour les jeunes du comté de Maricopa, division de la Cour supérieure de l'Etat d'Arizona. La juridicion du tribunal couvre une population totale d'environ 1,4 million de personnes, dont 428 475 ont moins de 18 ans.

Le système JOLTS a été conçu au départ pour contrôler des cas de jeunes délinquants et incorigibles, remplaçant un fichier manuel surchargé. Après trois années de planification, le système JOLTS a réalisé ses trois principaux objectifs:

1) fournir rapidement des renseignements précis sur les clients aux agents de probation et à d'autres organismes de l'appareil judiciaire;

2) fournir des rapports de gestion; et

3) accroître la capacité de recherche.

On a depuis ajouté au système de dépistage des cas, un système de circulation et l'établissement des calendriers des tribunaux.

La composante d'identification du système JOLTS utilise un système de classement phonétique qui contient le nom et le numéro de dossier de chaque jeune figurant à la base de données. L'identificateur primaire est le nom de famille du jeune et les identificateurs secondaires facultatifs sont le prénom, le sexe et la date de naissance. Tous les jeunes qui appartiennent au même groupe de noms sont considérés comme des appariements et apparaissent à l'écran. Le système JOLTS n'utilise aucune technique de cotation des probabilités comme telle, mais les appariements sont présentés en ordre d'exactitude pour aider le jugement. Un rapport mensuel de contrôle de la qualité permet d'identifier les duplications possibles.

Le système JOLTS fonctionne sur ordinateur HONEYWELL 66/40, tant par lots qu'en direct. Par suite d'une évaluation du système JOLTS effectuée récemment, on a recommandé la conversion à un matériel plus courant.

5.11 Le système pour les jeunes de Memphis - Application du système JISRA

Le système pour les jeunes de Memphis (Tennessee) est l'application la plus récente sur ordinateur principal du modèle JISRA (Juvenile Information System & Records Access). Il s'agit d'un projet du National Council of Juvenile and Family Court Judges, organisme américain. JISRA répond aux besoins de gestion, de déclarations extérieures et de recherche spéciale des tribunaux pour les jeunes; on le décrit comme un système informatique en direct de dépistage des événements et des personnes conçu pour le transfert.

Le système JISRA peut être transféré à tout matériel compatible avec les appareils IBM et à tout microprocesseur pouvant utiliser le systèmes OASIS. Puisque le système JISRA a été conçu pour divers environnements, il se présente sous forme d'une série de modules que l'on peut choisir ou réviser en fonction des besoins de chaque tribunal.

Les modules JISRA suivants, disponibles sur toutes les versions, peuvent fonctionner différemment dans chaque application:

1) Sécurité - La sécurité de l'opérateur est fonction à la fois de l'ordinateur principal et du micro-ordinateur. La sécurité du terminal n'est disponible que pour les ordinateurs principaux.

2) Edition - L'édition des champs de données se fait par table.

3) Jeunes	- Recueillir et conserver les données d'identification et les données environnementales concernant un jeune.
4) Recherche onomastique	- Pour établir s'il y a eu contact antérieur.
5) Sommaire du cas	- Donne les antécédents du jeune, y compris les dates, les renvois et leur aboutissement.
6) Renvois/admission	- Inscrit les renseignements de base pour chaque renvoi.
7) Calendrier	- Produit les calendriers du tribunal.
8) Audition/issue	- Inscrit les données sur l'audition, l'incident et l'issue.
9) Détention	- Inscrit les données sur les soin en résidence.
10) Statistiques	- Pour mesurer l'efficacité, indiquer les tendances et projeter les besoins.
11) Production de rapports	- Selon des formules préétablies.
12) Faux noms	- Plusieurs noms peuvent être retenus pour le même jeune.
13) Adulte	- Pour recueillir des renseignements sur les adultes associés avec le jeune.

14) Famille - Pour donner une liste descriptive de tous les membres de la famille.

15) Autres personnes/ témoins - Pour identifier les personnes associées à un renvoi ou à une plainte.

Le module de recherche onomastique de l'application sur ordinateur principal applique un algorithme phonétique au nom du jeune pour établir les contacts antérieurs. Le système produit une liste, en ordre alphabétique, de tous les jeunes figurant au fichier qui ont le même nom ou nom semblable (nom réel ou faux nom).

Le modèle JISRA n'utilise pas d'identificateurs secondaires. Les éléments de données qui apparaissent à l'écran pour aider à établir s'il y a eu contact antérieur sont le nom et le prénom, les faux noms, le sexe, la date de naissance et la race.

L'application de Memphis fonctionne sur ordinateur IBM modèle 4331.

5.12 Toronto Credits Limited - Application du secteur privé

Divers organismes ont besoin d'identifier des individus. Les besoins de l'appareil judiciaire et ceux des services sociaux sont semblables en ce qu'il faut suivre les individus dans le temps et d'un organisme à l'autre. Les buts et les méthodes de collecte des données des organismes du secteur privé peuvent être bien différents de ceux du secteur public, mais le problème d'identification est commun. Par exemple, les bureaux de crédit ont des fichiers importants de données concernant des individus et qui sont interrogés au moyen de divers critères de recherche. Les méthodes de collecte des données, les buts opérationnels de l'organisme et le genre de données sont particuliers

aux bureaux de crédit, mais les besoins fondamentaux de la technologie d'identification ne le sont pas.

Bon nombre d'organismes du secteur privé utilisent un code unique quelconque pour identifier les individus figurant au fichier (par exemple, un magasin à rayons aura un numéro de compte). Ce code est facilement disponible pour relier les individus parce qu'il est de l'intérêt de l'individu de le communiquer. D'autre part, un bureau de crédit ne peut pas toujours s'en remettre à un code unique de ce genre pour identifier les individus.

Toronto Credits Limited est un organisme privé qui a mis au point une technique d'identification qui pourrait être jugée adaptée aux besoins des systèmes d'information juridique. La société offre un service de rapports de crédit en direct aux bureaux membres, comme le Bureau de crédit sous le régime de la Loi sur la protection du consommateur. Ce service est offert aux clients contre rémunération. Un individu peut également demander communication de son dossier personnel.

Les données sur les individus sont recueillies à partir de diverses sources, notamment:

- les brefs;
- les jugements libératoires de faillite;
- les hypothèques mobilières;
- les demandes de crédit; et
- les clients.

Une base de données unique avec index à plusieurs niveaux utilisant une méthode d'accès direct (BDAM) a été mise au point par Toronto Credits pour mettre les données en mémoire et y avoir accès. Une fiche de consommateurs comprend des segments enchaînés, pour la

plupart de longueur variable, contenant des données relatives au nom, à l'identification personnelle, aux antécédents d'emploi, aux interrogations antérieures, à l'adresse, au commerce ou au compte ainsi que divers autres segments. La fiche du consommateur est mise en mémoire au moyen d'un numéro de contrôle qui désigne la fiche. On peut avoir accès aux données soit directement au moyen du numéro de contrôle ou par l'index en fournissant le nom et l'adresse et, de façon facultative, le prénom, le sexe, le nom du conjoint, l'âge ou le numéro d'assurance sociale. Un algorithme phonétique est appliqué au nom et à l'adresse avant la recherche de l'index. Une cote est accordée à chaque élément de données selon qu'il y a ou non appariement et les fiches présentant les quatre cotes les plus élevées dépassant un seuil prédéfini sont affichées.

Ce système fonctionne sur un grand ordinateur IBM modèle 3080A dans le cadre d'un bureau de service à temps partagé.

5.13 GIRLS - Application de liaison des dossiers

GIRLS (Generalized Iterative Record Linkage System) est un ensemble conçu par Statistique Canada pour déterminer des appariements probables de fiches qui n'ont pas de code unique d'identification. Le système sert présentement à des applications statistiques qui exigent l'appariement des fichiers et, sous sa forme actuelle, ne peut servir à des systèmes opérationnels exigeant l'appariement de fiches unique. Les concepts fondamentaux de liaison probabiliste des dossiers sont toutefois pertinents pour notre projet at bon numbre des routines de prétraitement des fichiers peuvent être appliquées à tout système d'information qui exige la préparation de champ avant l'appariement, qu'il s'agisse d'appariement de recherche des dossiers ou de liaison des dossiers.

Les système comprend plusieurs phases de traitement. Voici, en termes généraux, les fonctions de chaque phase:

1) Phase de prétraitement - Cette étape ne fait pas présentement partie intégrante du système, mais des routines sont à la disposition de l'utilisateur pour préparer les fichiers et qui peuvent comprendre la normalisation des champs du nom et de l'adresse et de la conversion des données.

2) Phase d'analyse - Le code et les tables de sources du programme sont générés d'après les définitions du fichier fournies par l'utilisateur et les règles de comparaison de champ.

3) Phase de génération des pondérations - D'après les comptes de fréquence, des pondérations sont générées et emmagasinées pour les divers champs de données visés par la liaison.

4) Phase de comparaison - Des liaisons possibles de paires de dossiers sont dterminées d'après des critères de comparaison, champ par champ, définis par l'utilisateur. Les dossiers sont groupés en "cases" ou blocs de dossiers selon les valeurs de certains champs clés et sont comparés physiquement uniquement dans ces

"cases". Les renseignements concernant les liaisons possibles sont emmagasinés dans la base de données du projet GIRLS.

5) Phase de pondération - Cette phase précise la pondération totale de la paire reliée.

6) Phase de liaison - D'après la pondération totale déterminée à la phase précédente et d'après les seuils définis par l'utilisateur, le système détermine si la paire constitue une liaison refusée, possible ou certaine.

7) Phase de groupe - Le système identifie des ensembles de dossiers qui sont reliés par des liaisons certaines ou possibles. Les "conflits" peuvent être résolus automatiquement par l'utilisateur. Les renseignements concernant les relations de groupe sont emmagasinés dans la base de données du projet.

8) Phase de rapport - L'information est produite à un niveau de détail concernant les liaisons, les pondérations et les groupes.

9) Phase d'information - Fournit des renseignements sur l'état actuel du traitement de liaison.

10) Phase de post-traitement - Cette phase, qui n'est pas une composante du système GIRLS, est la dernière opération de la liaison. Au moyen des renseignements fournis par les rapports GIRLS et des relations que comporte la base de données du projet, l'utilisateur peut effectuer des fonctions comme la constitution de dossiers composites ou la création de fichiers sommaires pour l'analyse des données.

Le système permet à l'utilisateur d'adjuster les paramètres en interaction jusqu'à ce qu'il obtienne les résultats optimums. La répétition non désirée d'étapes terminées avec succès n'est pas nécessaire. Le processus itératif est d'ordinaire exécuté en direct sur un échantillon de dossiers. Les ajustements des paramètres se font au terminal pour permettre à l'utilisateur d'expérimenter sur l'échantillon jusqu'à l'obtention de résultats optimums. A ce moment, l'utilisateur demande le traitement par lots pour obtenir des rapports imprimés qu'il pourra consulter pour établir les mesures à prendre après le traitement.

Un certain nombre de routines ont été mises au point pour la phase de prétraitement. Les routines sont écrites sous forme de procédures accompagnant un programme d'entraînement fourni par l'utilisateur et sont à la disposition de l'utilisateur pour effectuer certaines opérations précises de normalisation des champs et de conversion. Puisque ces routines ne sont pas liées à une application donnée, elles peuvent servir à tout système d'information qui doit identifier des individus. Des routines complètes servant à condenser et à normaliser tant les adresses (ASKGEN2, ENCODA) que les noms (NSKGEN5) ont été

mises au point pour effectuer la prétraitement pour le système de maintien du registre des entreprises, application qui utilise l'ensemble GIRLS. La routine d'adresse, en particulier, mérite d'être mentionnée car elle produit une clé de recherche très fiable qui minimise les erreurs de format et d'orthographe qui entravent d'ordinaire les recherches et les appariements fondés sur un adresse.

5.14

SUMMAIRE DES CARACTERISTIQUES

	Routine phonétique	Identificateurs facultatifs	Cotation	Base individuelle	Documentation disponible
Provincial Case File System de Colombie-Britannique	Mise au point interne	Oui	Non	Oui	Oui
Court Automated Project d'Alberta	D'après FIND	Oui	Non	Non	Oui
Corrections Management Information System d'Alberta	Russell Soundex	Oui	Non	Oui	Oui
Services correctionnels de Saskatchewan	Aucune	Oui	Non	Oui	Oui
Services correctionnels de l'Ontario	Mise au point interne	Oui	Non	Oui	Oui
Système d'information juridique du Nouveau Brunswick	D'après NYSIIS	Non	Non	Non	Oui
Services sociaux de Nouvelle-Ecosse	Mise au point	Oui	Non	Oui	Oui
Fichier des personnes (GRC)	FIND	Oui	Oui	Oui	Oui
Système judiciaire pour les jeunes de l'Utah	NYSIIS	Oui	Oui	Oui	Oui
Juvenile On-Line Tracking System du comté de Maricopa	Russell Soundex	Oui	Non	Oui	Oui
Système judiciaire pour les jeunes de Memphis	Mise au point interne (JISRA)	Non	Non	Oui	Non*
Toronto Credits Limited	Mise au point interne	Oui	Oui	Oui	Non
GIRLS			-S/O-		

* La documentation est disponible à l'achat de l'ensemble JISRA.

6. LES DÉTAILS DES TECHNOLOGIES

6.0 Introduction

Le présent chapitre donne une description plus détaillée des technique d'identification utilisées par cinq des systèmes examinés au chapitre 5, "Exposé sommaire des technologies".

Le but de la présente étude est de passer en revue les technologies actuelles d'identification comme point de départ des phases subséquentes de l'étude. C'est pourquoi les systèmes figurant à la présente section ne sont pas nécessairement les "meilleurs" systèmes intégrés d'information disponibles; ils représentent plutôt un ensemble des diverses démarches appliquées dans le domaine de la technologie d'identification.

6.1 Les services correctionnels de l'Ontario

Le Rapid Enquiry System (RES) permet l'interrogation en direct des données de base concernant les détenus actuels et passés du système correctionnel pour adultes. On interroge le RES pour établir s'il y a eu contact antérieur et pour produire des demandes de données complètes sur le profil du client du Adult Information System, à traitement par lots.

La base de données en direct RES est interrogée par les commis du groupe des dossiers des détenus chaque fois qu'un détenu entre dans une prison pour établir s'il a déjà eu contact avec le système. Les recherches peuvent se faire au moyen d'un de quatre types de clés:

1) non/sexe/date de naissance ou âge;
2) numéro de dossier principal (la clé primaire des systèmes RES et AIS);

3) numéro de l'établissement ou de la probation et de la libération conditionnelle; et
4) numéro de série des empreintes digitales.

La clé non/sexe/date de naissance peut être inscrite en entier ou partiellement, comme suit;

non/prénom/date de naissance
non/prénom;
non/prénom/sexe;
non/prénom/sexe/année de naissance;
non/prénom/sexe/mois de naissance/année de naissance et
non/prénom/âge (converti en année de naissance +3).

Une routine phonétique est appliquée au nom et au prénom pour dériver une clé qui est ensuite comparé à la clé équivalente de la base de données. L'opérateur doit inscrire le nom et le prénom au long pour assurer l'intégrité de la clé phonétique.

Le clé phonétique de base transforme toutes les lettres suivant la première lettre comme suit:

A E H I O U W Y	EN	ESPACES
C G J K Q X	EN	1
L	EN	2
M	EN	3
N	EN	4
B P	EN	5
R	EN	6
S Z	EN	7
D T	EN	8
F V	EN	9

Cette transformation donne la valeur codée de la lettre dans des circonstances normales. Il arrive que la lettre, dans le contexte des lettres environnantes, acquière un autre son ou devienne muette. Un ensemble complet de routines d'exception convertit les lettres dans les circonstances particulières. Ceci assure que les noms souvent mal orthographiés comme "JERRY" et "GERRY" soient appariés.

Après les routines de transformation et de conversion, la clé se présentera sous la forme suivante:

CODE DU NOM	A9999
CODE DU PRENOM	A9
CODE DU 2e PRENOM	A9 (il peut y avoir des espaces ou des zéros)

Les recherches partielles peuvent être effectuées à partir du seul nom, du nom et du sexe, du nom, du sexe et de l'année de naissance ou du nom, du sexe, de l'année et du mois de naissance. Si c'est l'âge qui est inscrit plutôt que la date de naissance, une variation allant jusqu'à trois années est permise.

La base de données en direct RES n'est pas mise à jour au terminal. Le fichier maître AIS est mis à jour et transmet les renseignements sur les détenus au RES au cours de la prochaine mise à jour.

6.2 Les services sociaux de Nouvelle-Ecosse

Le Client Index System de Nouvelle-Ecosse utilise une méthode simple pour déterminer les contacts antérieurs; cette méthode est adaptée aux besoins d'un système intégré d'information desservant un petit nombre de clients.

On peut entreprendre la recherche des contacts antérieurs en fournissant soit le numéro unique du client, s'il est connu, soit les identificateurs des caractéristiques personnelles. Comme solution de rechange, le numéro d'assurance sociale peut également être inscrit. Le reste de la section porte sur la façon de tenter un appariement au moyen des identificateurs des caractéristiques personnelles.

Le système exige le nom du client et, de façon facultative, le prénom ou les initiales, le sexe, l'âge ou la date de naissance.

La clé Soundex est constituée de la première lettre du nom, et les consonnes sont représentées par des chiffres. Les voyelles sont éliminées. En outre, la clé contient un code du sexe et l'année de naissance. La recherche du fichier se fait d'après cette clé, et les autres identificateurs fournis sont vérifiés.

NOM DE FAMILLE - Si le nom de famille est inscrit, le système énumère tous les clients portant ce nom. Si le nom de famille s'accompagne du prénom, tous les appariements de clients sont énumérés. Le nom de famille peut être inscrit avec la première lettre du prénom; tous les clients dont la première lettre du prénom et le nom de famille sont identiques sont énumérés. Enfin, jusqu'à quatre initiales de prénom, séparées par des virgules, peuvent être inscrites avec le même nom de famille. Tous les clients dont les initiales du prénom sont différentes, mais dont le nom de famille est identique sont énumérés.

PRENOM - Si le prénom est donné, seul un appariement exact est considéré comme une réussite pour cet élément. On n'effectue pas de recherches de nom partiel.

SEXE - Il faut inscrire "M" ou "F". La lettre "U" peut être inscrite pour effectuer une vérification des contacts antérieurs, mais ne peut être inscrite dans un fichier comme un identificateur primaire.

AGE OU
DATE DE
NAISSANCE - Le système permet d'effectuer une recherche sur trois années civiles une année avant l'année de naissance indiquée (ou calculée si l'âgeest inscrit) et une année après l'année de naissance. L'année et le mois précis peuvent également être inscrits.

S'il n'y a qu'une seule fiche appariée, un profil du client est affiché pour cet individu. Si plus d'un appariement possible est trouvé, une "liste abrégée" des individus est affichée. L'opérateur choisit alors les individus de cette "liste abrégée" pour un affichage plus complet des données du profil du client.

Les données du profil du client dans le système indique la participation de l'individu au programme. Chaque organisme conserve des fichiers détaillés portant sur l'activité de l'individu dans le programme.

6.3 La Gendarmerie royale du Canada (GRC)

La technique de recherche des fichiers en direct utilisée par le fichier des personnes de la GRC est un des systèmes de recherche onomastique les plus fiables présentement en usage. La routine phonétique, communément désignée sous le nom de code FIND (Fonetically Indexed Name Directory) a été adoptée par de nombreux systèmes y compris les systèmes d'immatriculation des provinces, la police d'Edmonton et les Archives nationales.

Le système de recherche onomastique de la GRC comprend quatre étapes fondamentales.

1) Déterminer le code primaire:

Une routine phonétique est appliquée au nom de famille pour déterminer le code primaire. Ce code indique le segment du fichier qui doit faire l'objet de la recherche.

2) Déterminer les codes XREF (renvoi):

D'autres routines phonétiques sont appliquées au nom de famille et les combinations possibles du code XREF et du code FIND désignent des segments différents du fichier pour la recherche.

3) Recherche par cotation:

Il y a recherche des segments de fichiers, des cotes étant attribuées aux éléments-clés de données selon qu'ils sont appariés, non appariés ou manquants et la cote totale de la fiche est calculée.

4) Liste:

Les appariements probables sont présentés en ordre descendant des cotes.

Le buts du codage des noms (étapes 1 et 2) est de segmenter le fichier et de placer les noms qui ont des resemblances phonétiques dans le même segment. Cette opération réduit le nombre de fiches qui doivent être parcourues matériellement pour déterminer un appariement. Le code FIND traduit phonétiquement le nom de famille

en un code à cinq caractères qui désigne le segment de noms qui présentent des ressemblances phonétiques. La première lettre du nom est transformée en l'un des caractères suivants:

B, D, E, F, J, K, L, M
N, P, R, S, T, U, W

et les autres lettres sont traduites en chiffres.

Il existe un problème commun à toutes les routines phonétiques: il vient un moment où de nouveaux raffinements du code en fonction de variations connues des noms dégradent le codage des variations existantes. Pour résoudre ce problème, un code XREF (renvoi) est généré pour tenir compte des variations. Les caractères du code primaire et du code XREF font ensuite l'objet d'un renvoi pour produire jusqu'à 64 combinaisons de codes qui désignent les segments du fichier qui doivent faire l'objet de la recherche.

Une fois les segments trouvés, les cotes des fiches sont générées en appliquant une technique de cotation de probabilité à chaque élément-clé de données. Les cotes sont calculées pour les appariements, les non appariements ou les éléments manquants au moyen des identificateurs suivants:

NOM DE FAMILLE*
DENTS-TRAITEES
DENTS-ABSENTES
DATE DE NAISSANCE*
AGE
PRENOMS*
LIEU DE NAISSANCE*
YEUX
TAILLE

POIDS
RACE
SEXE

L'astérisque indique les éléments qui obtiennent des points lors d'un appariement. Les autres perdent des points s'il n'y a pas appariement dans certaines limites.

Nous expliquons plus en détails la méthode de cotation des probabilités utilisée pour le nom de famille pour mieux faire connaître la façon dont certaines cotes sont attribuées. Il faut se rappeler que la cote de chaque élément de données est calculée de sorte qu'un seul écart n'exclue pas la fiche (la seule exception est le sexe, où un défaut d'appariement entraîne le rejet de la fiche).

Chaque fiche du segment faisant l'objet de la recherche présente automatiquement un nom phonétiquement semblable au nom de famille du sujet. Un appariement exact au sein du segment à l'égard d'un nom de famille peu fréquent reçoit une cote plus élévée que l'appariement exact d'un nom très répandu. A cette fin, la taille du plus grand segment est comparée au nombre d'occurrences du nom de famille du sujet. Ce rapport reçoit ensuite une cote d'après une échelle mise au point par Howard Newcombe et qui porte son nom.

Exemples:

RAPPORT	COTE
1-1	0
1-2	1
1-4	2
1-8	3
...	...
1-8192	13

Taille du plus grand segment	14 737 inscriptions
Nom de famille - Smith	11 449 inscriptions
Rapport	14 737 : 11 449 soit 1 à 1
Cote = 0	

Une fois les cotes assignées à chaque élément de données, la cote totale de la fiche est calculée. Les appariements probables sont alors présentés en order décroissant de probabilité. Pour aider le jugement, le système présente la cote de la fiche de même que la cote la plus élevée possible pour l'ensemble donné d'identificateurs.

6.4 Le système judiciaire pour les jeunes de l'Utah

Ce système utilise les routines de recherche onomastique d'un système autonome de traitement de l'identification appelé MICRONYM. MICRONYM, mis au point par SEARCH Group Inc. de Sacramento (Californie) est un système compltement indépendant pour la gestion et l'exploitation du traitement des documents et de la recherche onomastique dans un bureau d'identification d'Etat.

L'Utah a modifié le logical MICRONYM pour l'adapter à son ordinateur, mais le concept d'un index maître des noms distinct demeure inchangé. Le numéro unique du cas est la clé qui relie les fiches du même individu. Ce numéro est automatiquement augmenté d'une unité pour chaque nouveau sujet.

La première étape est de déterminer le groupe de fiches devant faire l'objet de la recherche. La méthode utilisée par MICRONYM a été élaborée en 1970 pour le New York State Information and Intelligence System (NYSIIS), et est connue communément sous le nom de méthode NYSIIS. Le processus comprend les étapes suivantes:

1. Si les premières lettres du nom sont:

   "MAC", alors changer ces lettres en "MCC"
   "KN", alors changer ces lettres en "NN"
   "K", alors changer cette lettre en "C"
   "PH", alors changer ces lettres en "FF"
   "PF", alors changer ces lettres en "FF"
   "SCH", alors changer ces lettres en "SSS"

2. Si les dernières lettres du nom sont:

   "EE", alors changer ces lettres en "Y"
   "IE", alors changer ces lettres en "Y"
   "DT" ou "RT" ou "RD" ou "NT", alors changer ces lettres on "D".

3. Le premier caractère du code phonétique est le premier caractère du nom d'entrée.

4. Dans les règles suivantes, un balayage des caractéres du nom est effectué. Ceci se décrit en termes d'une boucle de programme. Un pointeur sert à avoir accès à la position à l'étude dans le nom. L'étape 4 consiste à diriger ce pointeur sur le second caractère du nom.

5. Compte tenu de la position du pointeur, un seul des éconcés suivants peut être exécuté.

   Si le dernier caractère du nom a été exploré, passer à la règle 7.

   Si la position actuelle est une voyelle (AEIOU):

      alors si elle est égale à "EV", changer en "AF"
      autrement changer la position actuelle en "A".

Si la position actuelle est la lettre:

"Q", alors changer la lettre en "G"
"Z", alors changer la lettre en "S"
"M", alors changer la lettre en "N"

Si la position actuelle est la lettre "K":

alors si la lettre suivante est "N", remplacer la position actuelle par "N"
autrement, remplacer la position actuelle par "C".

Si la position actuelle est la chaîne de lettres

"SCH", alors remplacer la chaîne par "SSS"
"PH", alors remplacer la chaîne par "FF".

Si la position actuelle est la lettre "H" et que la lettre suivante ou précédente n'est pas une voyelle (AEIOU), alors remplacer la position actuelle par la lettre précédente.

Si la position actuelle est la lettre "W" et si la lettre précédente est une voyelle, alors placer la position actuelle avec la position précédente.

Si la position actuelle est un espace ou "-", alors remplacer la position actuelle par la position précédente.

Si aucune de ces règles ne s'applique, conserver la valeur de la lettre de la position actuelle.

6. Si la lettre de la position actuelle est ègale à la dernière lettre placée dans le code, placer le pointeur à la lettre suivante et passer à l'étape 5.

   Le prochain caractère du code est la lettre de la position actuelle.

   Déplacer le pointeur d'un cran pour arriver à la lettre suivante.

   Passer à l'étape 5.

7. Si le dernier caractère du code est la lettre "S", la retrancher. Passer à l'étape 8.

8. Si les deux derniers caractères du code sont les lettres "AY", alors les remplacer par le caractère "Y". Passer à l'étape 9.

9. Si le dernier caractère du code est la lettre "A", la retrancher. Le code qui résulte de ces opérations est une chaîne de caractères de longueur variable. Il est possible de créer un code vide où aucune lettre du nom n'est conservée sous quelque forme que ce soit. Ce code vide serait créé par un nom comportant uniquement des voyelles et se terminant en "S". Si un code vide est généré, les trois premières lettres du nom servent de code.

10. Placer la première initiale en position 10 du code de dix caractères, et tronquer le cas échéant à dix caractères.

Une fois le groupe de noms déterminé, chaque fiche est explorée et des cotes sont attribuées aux éléments de données suivants:

Nom de famille
Prénom

Première initiale
Deuxième prénom
Deuxième initiale
Sexe
Race
Mois de naissance
Jour de naissance
Année de naissance
Type d'empreinte digitale
Numéro de sécurité sociale

Dans le système MICRONYM, il y a une cote d'appariement et une cote de non appariement pour chaque élément de donnée. Il n'y a aucune évaluation de la "gravité" de l'erreur". Un élément manquant obtient la cote zéro. Un modèle mathématique appelé le théorème de Bayes a été choisi pour déterminer les cotes quantitatives de chaque élément et la cote globale de la fiche. Si un seuil prédéterminé n'est pas atteint, la fiche est considérée un échec.

Les résultats de la recherche sont présentés à l'opérateur qui informe le système des résultats de la vérification de la recherche. Une nouvelle fiche maître peut être créée, des fiches de faux noms peuvent être créées ou les parties vides d'une fiche existante peuvent être remplacées par de nouveaux renseignements.

6.5 Le Juvenile On-Line Tracking System (JOLTS) du comté de Maricopa

Les fichiers d'identification du réseau JOLTS sont organisés par un système phonétique très semblable au premier code Russell Soundex. Chaque jeune a une fiche contenant des données détaillées d'identification qui servent de renvoi au numéro de fichier assigné à son nom. Ces dossiers détaillés sont regroupés en fiches maîtresses Soundex d'après les caractéristiques phonétiques des noms plutôt que

d'après la graphie. La base de données JOLTS comprend environ 3 200 fiches maîtresses Soundex et 45 574 fiches détaillées Soundex.

Les éléments de données utilisés pour déclencher une recherche sont:

Nom de famille (obligatoire)
Prénon (facultatif)
Sexe (facultatif)
Date de naissance (facultative)

Le système Soundex convertit le nom de famille en un code de quatre caractères au moyen des critères suivants:

1) Le premier caractère est le premier caractère du nom de famille

2) Les deuxième, troisième et quatrième caractéres sont remis à zéro puis calculés comme suite:

a) A, E, I, O, U, Y, H, W	sont négaligés
b) B, P, F, V	sont convertis à 1
c) C, G, J, L, G, S, X, Z	sont convertis à 2
d) D, T	sont convertis à 3
e) L	est converti à 4
f) M, N	sont convertis à 5
g) K	est converti à 6

Ce code est alors utilisé directement pour retrouver une fiche maîtresse Soundex. Les fiches détaillées sont alors retrouvées en ordre et affichées à l'écran dans l'ordre suivant:

1) Appariement exact du nom de famille, du prénom, du sexe et de la date de naissance.

2) Appariement exact du nom de famille, du prénom et du sexe.

3) Appariement exact du nom de famille, de la première initiale et du sexe.

4) Appariement exact des quatre premiers caractères du nom de famille et du sexe.

5) Toutes les autres fiches qui consituent des détails pour la fiche maîtresse Soundex.

L'affichage des résultats de la recherche comprend les champs de données suivants:

Nom de famille
Prénom
Sexe
Date de naissance
Race
Type de cas
Faux nom
Prénom de la mère
Numéro de dossier

Des rapports de contrôle de la qualité sont effectués chaque mois pour indiquer les duplications possibles (le même jeune ayant plus d'un numéro de dossier).

ANNEXE A - LISTE DES CONTACTS

D. Anderson	Ministère des Services sociaux de la Saskatchewan
O. Anderson	Solliciteur général de l'Alberta
J. Barrett	Maricopa County Juvenile Court, Phoenix, Arizona
R. Beattie	Centre canadien de la statistique juridique
L. Bennett	Gendarmerie royale du Canada
D. Brown	Centre canadien de la statistique juridique
S. Burbridge	Solliciteur général du Canada
J. Burns	Ministère de la Justice du Nouveau-Brunswick
R. Carter	Solliciteur général de l'Alberta
W. Cook	Toronto Credits Limited
D. Clark	Ministère de la Justice du Nouveau-Brunswick
T. Conrad	Ministère des Services sociaux de la Nouvelle-Ecosse
F. D'Argis	Ministère du Procureur général de la Colombie-Britannique

ANNEXE A (suite)

A. Desmarais	Statistique Canada
B. Doherty	Juvenile Courts, Memphis, Tennessee, USA, Juvenile Information Systems & Records Access (JISRA)
R. Ferguson	Ministère de la Justice de la Saskatchewan
J. Fizzera	Ministère des Services sociaux de la Nouvelle-Ecosse
R. Girard	Ministère de la Justice, Montréal Québec
B. Grainger	Centre canadien de la statistique juridique
T. Hill	Statistique Canada
D. Jacobson	Ministère de la Justice de la Saskatchewan
F. Kaptein	Systems West Limited, Edmonton, Alberta
I. Kowaliw	Toronto Credits Limited
N. Lee	Centre canadien de la statistique juridique
F. Letkeman	Ministère du Procureur général de la Colombie-Britannique

ANNEXE A (suite)

J. Linton	Ministère des Services communautaires et sociaux de l'Ontario
J. Maginski	Reader's Digest, Montréal, Quebec
D. Martin	Services correctionnels et communautaires du Manitoba
C. McCartney	Ministère des Services sociaux et communautaires de l'Ontario
Insepcteur J. McDougall	Gendarmerie royale du Canada
M. Morin	Bureau de crédit d'Ottawa Limitée
B. Muhly	Procureur général de l'Alberta
J. O'Gorman	Ministère des Services correctionnels de l'Ontario
R. Ouellet	Ministère de la Justice du Québec
M. Phillips	Utah Juvenile Courts, Salt Lake City, Utah, USA
J. Pirozzi	National Council of Juvenile and Family Court Judges, Reno, Nevada, USA
L. Radulescu	Service de police d'Ottawa
E. Rathbun	National Crime Information Centre, New York, USA

ANNEXE A (suite)

J.A. Schriel	Santé et Bien-être social Canada
L. Shapiro	Ministère des Services communautaires et sociaux de l'Ontario
S. de Silva	Centre canadien de la statistique juridique
Chef adjoint S. Stevens	Service de police d'Edmonton
T. Strickland	Ministère des Services sociaux de la Saskatchewan
J. Townesend	Solliciteur général du Canada
P. Trommelen	Ministère des Services sociaux de l'Alberta
R. Warner	Services correctionnels et communautaires du Manitoba
T. Wilson	SEARCH Group Inc., Sacramento, California, USA